UNIVERSITY OF EDINBURGH

College of Science and Engineering
School of Chemistry

New Tools for High Throughput Chemistry and Biology

By
Luciano Paulo Rosa Galveia

Doctor of Philosophy
February 2011

UNIVERSITY OF EDINBURGH

ABSTRACT

Faculty of Engineering and Sciences

School of Chemistry

Doctor of Philosophy

New Tools for High-Throughput Chemistry and Biology

by Luciano Paulo Rosa Galveia

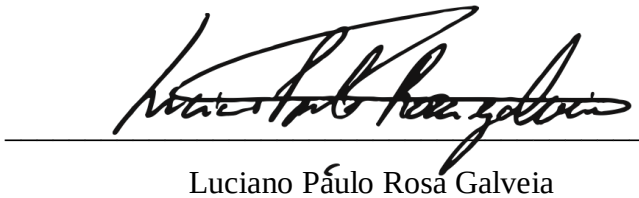
New strategies for the high-throughput determination of protease substrate specificity using a dual colour microarray based format with a small library of FRET-based peptides were developed. Integral to this process was the development of an in-house hydrophobic microarray platform and the application of microarrays of microwells. Overall this allowed the rapid and quantitative analysis of a range of peptides with three proteases, subtilisin, chymopapain and thermolysin.

An immobilised tetrafluoroarylsulfonyl chloride linker was developed to be used as a traceless linker for microwave-assisted palladium chemistry, namely transfer hydrogenation and deuteration, and Heck, Stille, Suzuki-Miyaura and Buchwald-Hartwig cross-coupling reactions. Palladium-mediated protocols under microwave irradiation were performed, allowing transfer hydrogenation and deuteation, and Suzuki-Miyaura and Heck cross-couplings.

DECLARATION

I declare that:

- This thesis has been composed by myself
- The work presented in this thesis is all my own
- No part of this thesis has been previously submitted at these or any other University for any other degree or professional qualification.



Luciano Paulo Rosa Galveia

10th of July of 2011

PREFACE

The research described in this thesis was carried out under the supervision of Prof. Mark Bradley at the University of Southampton (Feb. 2004 – Jan. 2005) and the University of Edinburgh (Feb. 2005 – February 2011).

ACKNOWLEDGMENTS

I would like firstly to thank my supervisor, Prof. Mark Bradley for all his teachings, for the enthusiasm, for his availability, trust and support during these years. I would also like to thank all the members of Prof. Bradley's research group, especially the members of the Combinatorial Centre of Excellence: Eric, Alessandra, Gianluca, Christophe and Delphine. In particular, I would like to thank all my friends for their sensibility, constant support e delicious laughs through these years.

I am also very grateful to Juanjo for all his help and time on the peptide constructs project, to Ferdous for the demonstrations in how to prepare the microwell slides, and to Jose for is teachings in the use of the software tools for design of experiments.

Finally, I would like to thank my family for their continual support all along my studies and during my PhD, especially my brother Helder Galveia.

Dedicated to my parents for everything.

TABLE OF CONTENT

Chapter Three: Every long walk starts with a single step…and then you want to run 83

Abbreviations

Aminoacids

A/Ala	Alanine
R/Arg	Arginine
N/Asn	Asparagine
D/Asp	Aspartic acid
C/Cys	Cysteine
Q/Gln	Glutamine
E/Glu	Glutamic acid
H/His	Histidine
I/Ile	Isoleucine
L/Leu	Leucine
K/Lys	Lysine
M/Met	Methionine
F/Phe	Phenylalanine
P/Pro	Proline
S/Ser	Serine
T/Thr	Threonine
W/Trp	Tryptophan
Y/Tyr	Tyrosine
V/Val	Valine

General

Ac	Acetyl
ACC	7-amino-4-caboxymethylcoumarin
AMC	7-amino-4-methylcoumarin
AMP	Amino methyl polystyrene
Anal.	Elemental Analysis
ANS	1-anilinonaphtalene-8sulphonic acid
Bn	Benzyl
Boc	*tert*-Butoxycarbonyl
br	Broad
Bz	Benzoyl
CCD	Charge-Coupled Device
*c*Hex	*Cyclo*hexyl
Chym	Chymotrypsin
d	doublet

Da	Dalton
DABCO	Bi*cyclo*[2,2,2]-1,4-diazaoctane
DANSen	5-dimethylaminonaphtalene-1-(*N*-2-aminoethyl)sulphonamide
DCC	Di*cyclo*hexylcarbodiimide
DCE	Dichloroethane
DCM	Dichloromethane
DCU	Di*cyclo*hexylurea
Dde	*N*-[1-(4,4-dimethyl-2,6-dioxo*cyclo*hex-1-ylidene)ethyl]
DEAD	Diethylazodicarboxylate
DEPT	Distortionless Enhanced Polarization Transfer
DIAD	Diisopropylazodicarboxylate
DIPEA	*N, N*-Di*iso*propylethylamine
DMA	Dimethylacetamide
DMAP	4-Dimethylaminopyridine
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DoE	Design of experiments
dppp	1,3-bis(diphenylphosphino)propane
EDANS	5-((2-aminoethyl)amino)naphthalene-1-sulphonic acid
ELSD	Evaporative Light Scattering Detector
ES	Electrospray
FAM	6-carboxyfluorescein
Fmoc	9-Fluorenylmethoxycarbonyl
FTIR	Fourier Transform Infrared
HIV	Human immunodeficiency virus
HMPA	Hexamethylphosphoramide
HPLC	High Performance Liquid Chromatography
HR	High Resolution
HTS	High-Throughput-Screening
*i*BuOH	*Iso*butanol
LCMS	Liquid Chromatography – Mass Spectrometry
m	Multiplet (NMR) or medium (IR)
MAS	Magic Angle Spinning
Bp	Boiling Point
Mp	Melting Point
MS	Mass Spectrometry
MW	Molecular Weight
NMM	*N*-methyl-morpholine
NMP	*N*-Methyl-2-Pyrrolidone
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffered Saline

P.E.	Petroleum Ether
PEG	Polyethyleneglycol
Ph	Phenyl
PMA	Phosphomolybdic acid
PNA	Peptide nucleic acid
PS	Polystyrene
RNA	Ribonucleic acid
s	Singlet (NMR) or strong (IR)
SDS	Sodium *n*-Dodecyl Sulfate
SPOS	Solid Phase Organic Synthesis
SPPS	Solid Phase Peptide Synthesis
SPE	Solid Phase Extractor
TAP	Tris(dimethylamino)phosphine
TBAF	Tetra-*n*-butylammonium fluoride
TBS	*Tert*-butyl-dimethyl-silyl
TEA	Triethylamine
TES	Triethylsilane
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIS	Tri*iso*propylsilane
TLC	Thin Layer Chromatography
TMAD	*N*, *N*, *N´*,*N´*-Tetramethylazodicarbonamide
TMU	Tetramethylurea
TPP	Triphenylphosphine
t_R	Retention time
UV	Ultraviolet
w	weak
Z	Benzyloxycarbonyl
μω	microwave

Coupling Reagents and Additives

CDI	Carbonyldiimidazole
DIC	Diisopropylcarbodiimide
EDC	1-Ethyl-3-(3-Dimethylaminopropyl)carbodiimide
HATU	*O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	*O*-(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOAt	1-Hydroxy-7-Azabenzotriazole
HOBt	1-Hydroxy-1*H*-Benzotriazole
PyBOP	Benzotriazol-1-yloxytri(pyrrolidino)-phosphonium hexafluorophosphate

PyBrop	Bromotri(pyrrolidino)phosphonium hexafluorophosphate
TATU	*O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate
TBFH	*N*, *N*, *N*', *N*'-Tetramethylbromoformamidinium-hexafluorophosphate
TBTU	*O*-benzotriazol-1-yl-1,1,3,3-tetramethyluronium tetrafluoroborate
TCTU	1 H-Benzotriazolium, 1- [bis (dimethylamino) methylene] -
TFFH	Tetramethylfluoroformamidinium hexafluorophosphate

Chapter One

High-Throughput Proteases Assays

The purpose of this chapter is to briefly describe proteases, report on many of the different high-throughput techniques used to determine protease substrate specificity and then introduce a novel approach for the analysis of protease substrate specificity.

1. Introduction

In the enigmatic web of life, biomolecules interact and connect in the most unexpected and unpredictable ways. Dissection of these interactions at the molecular level, both spatially and temporally, has been complex and difficult, and has lead to a number of different research fields such as proteomics and genomics. Proteomics[1] for example focuses on the study of proteins, post-translational modifications, interactions with other proteins, and their temporal expression. Their sequence, and structure holds the key to a better understanding of protein and possibly cellular function under both normal and diseased states.[2] In the case of genomics[3] for example, the analysis of the human genome, suggests the presence of 30,000-40,000 genes[3] and presents scientists with the daunting scenario of more than 100,000 different proteins (due to post-translational modification) being expressed only in the human species, which increases in complexity when protein-protein interactions have to be considered. Consequently, there have been rapid advances in the techniques and methods capable of large scale proteomic studies. From the various families of proteins that have been identified, enzymes are among the most interesting to investigate because of their unique catalytic specificity and efficiency. In the family of enzymes, proteases occupy one of the main roles in the regulation of life (**figure 1.1**).

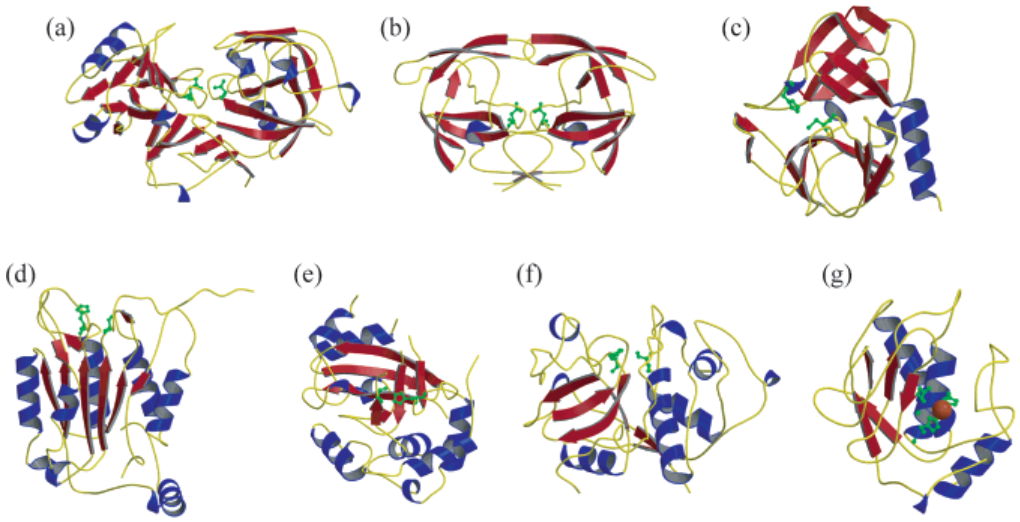


Figure 1.1: Seven common structural folds found among proteases. From top left are secondary structure representations of the (a) pepsin-like aspartic protease *β*-barrel (1pso), (b) retropepsin aspartic protease *β*-barrel (7hvp), (c) trypsin-like fold *β* -barrel (1cqq), (d) subtilisin-like and caspase-like *αβα* sandwich or Rossman fold (1ice), (e) herpes virus serine protease *αβ* barrel (1cmv), (f) papain-like cysteine protease *αβ* complex fold (1pap), (g) thermolysin-like catalytic domain (1mmq). Catalytic residues shown in green; the zinc atom is shown as an orange ball (thermolysin). Figures were generated using Molscript 2.0[4] and Raster3D v2.3[5].

1.1. Proteases: an important family of enzymes

Proteases are a class of enzymes that occur naturally in all organisms and constitute 1-5% of the gene content.[3] Until recently, proteases were considered to be merely protein-degrading enzymes, but this view has changed and proteases are now seen as important signalling molecules that are involved in many vital, often cascade processes.[6]

Proteases catalyse the hydrolysis of amide bonds between two amino acids, often with high specificity regarding the sequence of these amino acids. Proteases specifically cleave protein substrates either from the terminus N or C (exopeptidases) and/or from the middle of the molecule (endopeptidases). A specific nomenclature is given to a peptide that incorporates a scissile amide bond: each amino acid of the sequence is denominated by the letter P, with a number being attributed to it, depending on its proximity to the cleavage centre. Depending on whether it is located at the N or C section of the cleavable bond it is designated “prime” (C part). A similar principle is used with the letter S for the enzyme subpockets accommodating

the amino acids of the substrate (**figure 1.2**). This specificity of a protease not only relates to the two amino acids involved directly in the cleavage site (P_1 and P_1') but also the next positions (P_2, P_2', P_3, P_3', etc.) which makes/demonstrates a protease's high selectivity for a specific sequence.

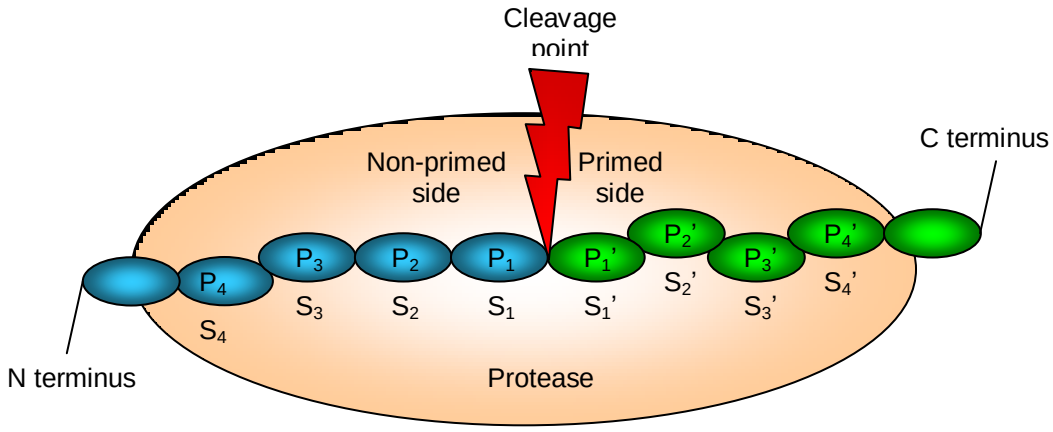


Figure 1.2: Schematic representation of substrate/protease binding.

All proteases must overcome three obstacles when hydrolyzing a peptide bond: (a) amide bonds are very stable due to electron donation from the nitrogen to the carbonyl (for comparison, a simple alkyl ester is ≈3000× more reactive than an amide bond, while a *p*-nitrophenyl ester is 300,000× more reactive[7]), proteases must activate an amide bond via the interaction of the carbonyl oxygen with a general acid, and may also distort the peptide bond to disrupt resonance stabilization.

Proteases are classified according to their activities and the functional groups involved in catalysis. There are currently six classes of proteases[8]: serine proteases, threonine proteases, cysteine proteases, metalloproteases, aspartic acid, and glutamic acid proteases. However, due to the small number of known threonine and glutamic acid proteases (and the fact that they were only relatively recently discovered, and their similarity to other groups), proteases are usually divided into four major classes: aspartic, serine, cysteine, and metallo-proteases. Mechanistically, these four classes can be divided into two groups of proteases. Since water is a poor nucleophile; proteases always activate water, usually via a general base to enhance reactivity. Since amines are poor leaving groups; proteases must provide some device to protonate the amine prior to expulsion. One group employs an enzyme-bound, activated water molecule as the nucleophile which attacks the amide carbonyl of the

scissile bond. This water can either be activated by ligation to a cation, or by binding in a small cleft defined by two aspartic acid residues (the aspartate proteases), or by binding to a glutamine and glutamate residues (the glutamic acid proteases). This class of enzymes catalyze direct amide hydrolysis by water. The second group of enzymes activate a nucleophilic group of an amino acid residue to initiate amide hydrolysis. To date, the hydroxyl group of a serine or threonine or the thiol group of cysteine, have been identified as enzyme-activated nucleophiles. These classes of enzymes catalyze amide hydrolysis directly. The nucleophilic atom of the enzyme attacks the amide carbonyl of the scissile bond and breaks the C-N bond, to form an acyl enzyme intermediate (either an ester or a thioester). This acyl enzyme intermediate is then hydrolyzed by water to complete the hydrolysis process.

1.1.1. The aspartic proteases

The aspartic proteases are endopeptidases that have catalytically important carboxylates and a low pH optima (except chymosin, whose activity extends to neutral pH's). Aspartic proteases have one of the longest histories in enzymology. Rennet (and later chymosin) have been used in the making of cheese for thousands of years.[9] However, pepsin is usually considered the first enzyme in this family, as it was the first to be recognized as an active component (i.e., the observation that gastric juice dissolves meat, in 1783) and given a name (in 1825), and was only the second to be extracted (in 1836) or crystallized (in 1930).[10] The second enzyme in the family is nepenthesin, detected in insect eating plants by Darwin[11] (in 1875), thereby making this the oldest family. Extracts of pepsin have to be acidified to regain full activity, with different acids giving different results. Sörensen[12] noted that if activities were plotted against hydrogen ion concentrations, results were similar. To solve his scaling problem, he employed logarithmic abscissas and introduced pH (in 1909). The identifying characteristic of what were the acid proteinases is their wide bell-shaped pH profiles with acidic optima, and this fundamental characteristic has never been accounted for, despite being featured in this, the primal report of pH-dependent enzymatic activity. For nearly a century, two acidic groups were thought to be involved in the catalytic activity of pepsin, one protonated and one not, with

p*K*s approaching extremes of 4 and 1, respectively. The first was obviously a carboxyl, but the second seemed much too acidic to be so as well. When a phosphoric acid residue was found in porcine pepsin half century ago, it became a candidate for the low p*K*, but dephosphorylation had no effect on activity. When both groups were deemed carboxyls, different environments seemed necessary, so one was attributed to the hydrophobic active site of the enzyme and the other to peptide substrates. But esterification of carboxyl groups of substrates had no effect either, thereby leaving the origin of the low p*K* unresolved for another 30 years.

That long uncertainty reaches across history is without precedent and, moreover, is accompanied by other seemingly unrelated uncertainties. Frustration peaked with transpeptidation reactions involving both halves of peptide substrates, implying covalent acyl and amino enzyme intermediates, whose parallel existence within a single enzymatic mechanism seemed unlikely.[13] Driven by the clinical significance of the HIV proteases and the intense search for mechanism-based inhibitors, the report of a low-barrier hydrogen bond (LBHB) determined by computational methods allowed for the various uncertainties to be resolved within a unifying hypothesis.[14] Previous proposals for chemical mechanisms have one carboxyl group, acting as a general acid to donate its proton to the substrate scissile carbonyl, and the other carboxyl unprotonated, acting as a general base to accept a proton from a water molecule.[15] But the LBHB ties up the carboxyl proton and rules these proposals out, solving this mystery in enzymology. Hunkapiller *et al.*[16] observed bell-shaped pH profiles for hydrolysis of *N*-trifluroacetyl-L-phenylalanine by pepsin that narrowed in D_2O; i.e., the lower p*K* shifted right but the upper pK shifted left. Normally, D_2O shifts all p*K*s to the right by about half a unit because of the small fractionation factor of H_3O^+.[17] Re-analysis of the pH profiles of k_{cat} from Hunkapiller are presented in **figure 1.3**.

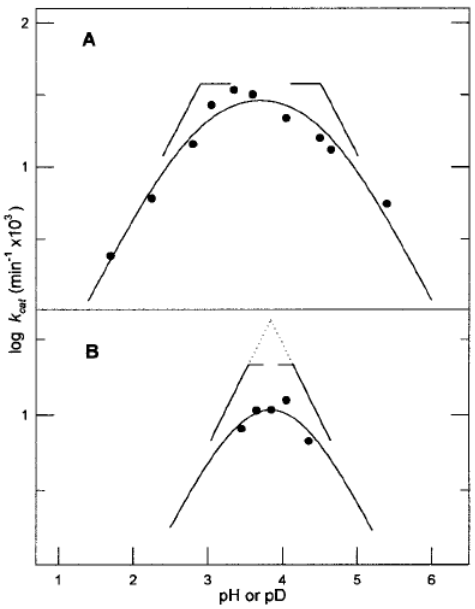
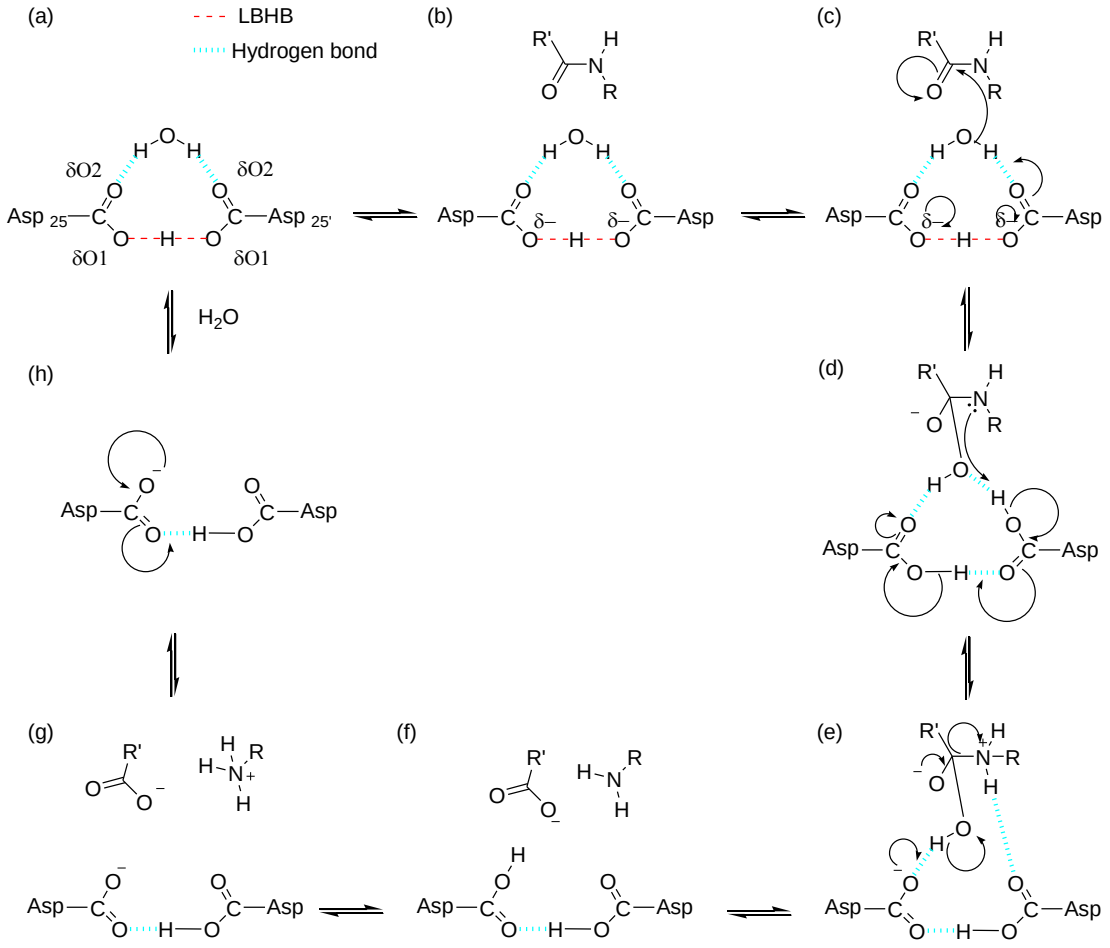


Figure 1.3: Kinetics of pepsin acting on *N*-trifluroacetyl-L-phenylalanine as a function of pH (A) and pD (B). The fitted pH-independent kinetic constants are k_{cat} = 0.038 ± 0.006 min^{-1} in H_2O and k_{cat} = 0.022 ± 0.002 min^{-1} in D_2O. The dotted line indicates the range of carboxyl p*K*s consistent with the fitted line in D_2O. (Similar shapes of profiles were obtained using values of k_{cat}/K_m)[16] (reproduced with permission).

The fit to the BELL pH equation in H_2O converged with p*K*s of 2.9 ± 0.1 and 4.5 ± 0.1, while those in D_2O converged to the default separation of only 0.6 unit, the minimum statistically possible. The fitted curve in **figure 1.3(B)** is the same for separated p*K*s of 3.5 and 4.1, or identical p*K*s of 3.8 ±0.1, or anything in between. LBHBs cause opposite shifts in the p*K*s of the conjugate bases participating in the bond. Both protonation and deprotonation of the LBHB complex are more difficult, which shifts apparent p*K*s outward, much more so in H_2O than in D_2O.[18] Hunkapiller *et al.*[16] estimated a $^{D2O}k_{cat} \approx 3$ by comparing the maximal values of their fitted curves. However, the more appropriate comparison is between the pH-independent maxima[17], represented by the horizontal bars in **figure 1.3**. That ratio yields $^{D2O}k_{cat}$ = 1.8 ± 0.4 for p*K*s separated by 0.6 unit, and lower values if the fractionation factor is greater than Φ = 0.3. Hence, the actual *kinetic* isotope effect is modest, and most of the rate difference has a *thermodynamic* origin: weakening of the LBHB causes a large decrease in enzyme poised for catalysis. The narrowness of these pH profiles in D_2O suggests that this pH-dependent step is fully expressed. In contrast, pH profiles of HIV-1 protease show both p*K*s of k_{cat}/K_m farther apart and both shifted to the right in D_2O, with only the lower p*K* expressed in profiles of k_{cat}.[19] These contrasts require that the pH-dependent step being expressed in this HIV-1 protease study is different from the one in the pepsin study, that it comes later in the kinetic mechanism than the loss of the LBHB, and that it is not fully expressed which causes apparent p*K*s to be

shifted outward even farther. Piana and Carloni[14] performed *ab initio* molecular dynamics simulations on HIV-1 protease, focusing on the conformational flexibility of the active site carboxyls, Asp 25 and Asp 25'. They found the monoprotonated form to be the most stable, with a LBHB between the Oδ1 atoms that compensates for their negative charge repulsion and holds them within an interatomic distance of 2.5 ± 0.1 Å. Moreover, as shown in **scheme 1.1**, this LBHB along with peptide dipoles holds the carboxyls in a coplanar conformation with a water molecule between-and hydrogen bonded to-both Oδ2 atoms. The resulting symmetrical 10-atom cyclized structure thus provides a scaffold to impart proximity, orientation, and nucleophilicity to the water molecule (**scheme 1.1(a)**). In **scheme 1.1(a)** not only does the LBHB hold the carboxyls in a coplanar configuration, but it also has the water molecule held in position by hydrogen bonds to oxygens opposite the LBHB. The water must be near the scissile bond, and that places the LBHB proton out of the reach of the substrate. Starting with this cyclized structure as enzyme form, a proposed chemical mechanism is described.[19]



Scheme 1.1: Aspartic proteases general catalytic mechanism.[19]

After substrate binding and flap closing (**scheme 1.1(b)**), a movement of electrons (**scheme 1.1(c-d)**) within the cycle and extending through the scissile carbonyl moves two protons and generates a tetrahedral intermediate bound to a diprotonated form of the enzyme. The next step (**scheme 1.1(d)**) moves two protons and generates the zwitterion intermediate bound to a monoprotonated form of enzyme. Finally, collapse of the zwitterion (**scheme 1.1(e)**) cleaves the scissile bond, destroys the coplanarity of the carboxyls, and leaves the enzyme in the planar form. That completes the chemistry with regard to the substrate, but not for the enzyme. Flap opening and product dissociation gives the free enzyme in the diprotonated form (**scheme 1.1(f)**), the form that catalyzes the release reaction. To complete a turnover,

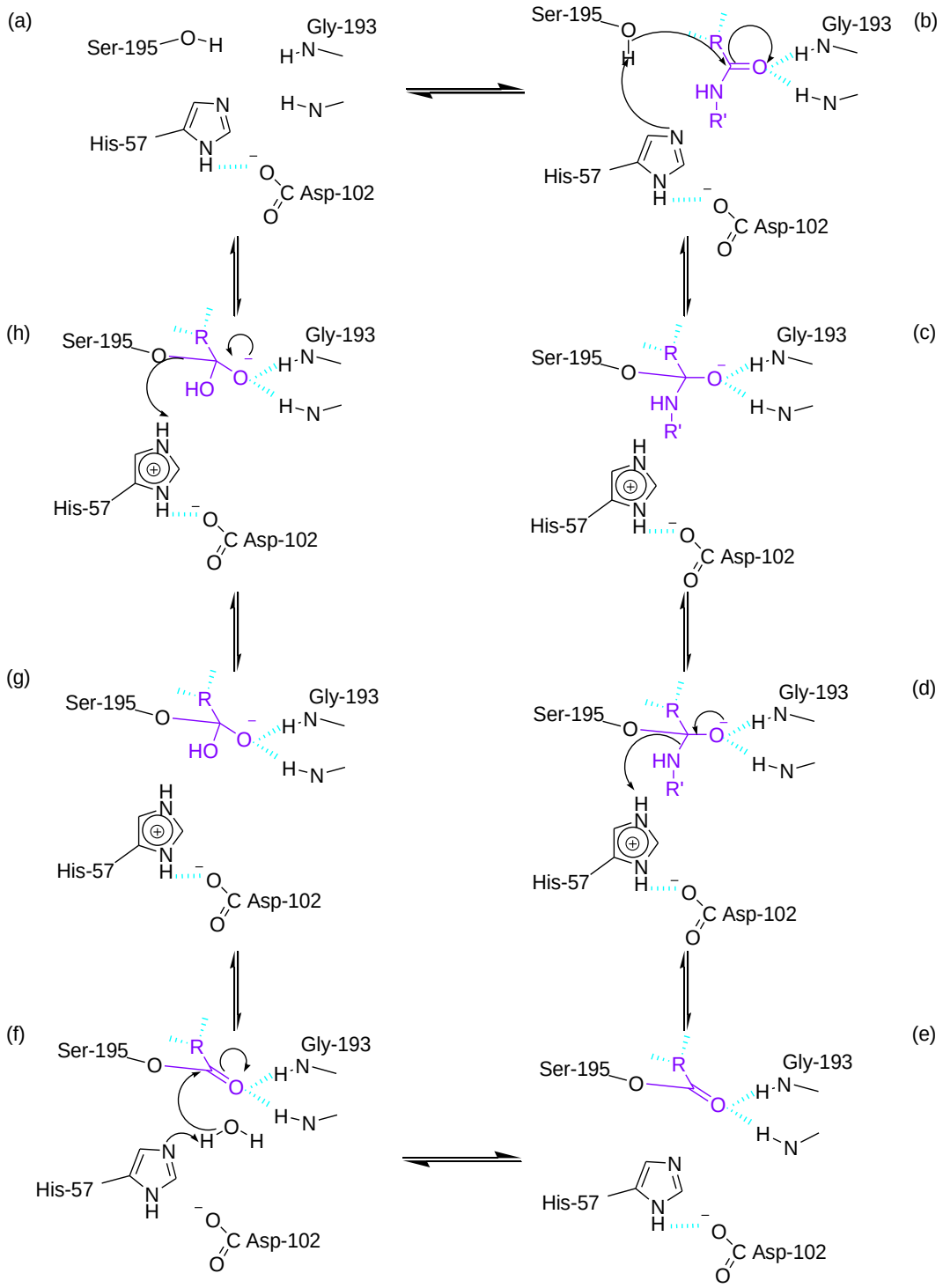
this form must be deprotonated (**scheme 1.1(g)**), rehydrated (**scheme 1.1(h)**), and allowed to restructure the 10-atom cyclic structure with the LBHB (**scheme 1.1(a)**).

1.1.2. The serine proteases

The serine proteases are peptidases that have a reactive serine residue and a pH optima around neutrality. Almost one-third of all proteases can be classified as serine proteases, named often for the nucleophilic serine residue at the active site. This mechanistic class was originally distinguished by the presence of the so-called Asp-His-Ser "charge relay" system or "catalytic triad".[20] The triad can be found in at least four different structural contexts, indicating that this catalytic machinery has evolved on at least four separate occasions.[21] These four "clans" of serine proteases are typified by chymotrypsin, subtilisin, carboxypeptidase Y, and Clp protease.[8] More recently, serine proteases with novel catalytic triads and dyads have been discovered, including Ser-His-Glu, Ser-Lys/His and His-Ser-His.[21]

Serine proteases have been found in vertebrates, bacteria, and also viruses. Examples of vertebrate members of this family include chymotrypsin, trypsin, thrombin, elastase, and kallikrein[22]. Bacterial proteases of this family include *Streptomyces griseus* protease A (SGPA), SGPB, and R-lytic protease, and viral proteases of this family include alphavirus capsid proteins and hepatitis C virus NS3 protease.[23]

Serine proteases perform their tasks very efficiently: the rates of peptide hydrolysis by serine proteases are $\approx 10^{10}$-fold greater than the uncatalyzed reactions. Obviously, the mechanisms of catalysis are not confined to peptide hydrolysis; serine proteases readily hydrolyze other acyl compounds, including anilides, esters, and thioesters. In the generally accepted mechanism[24-26] for chymotrypsin-like serine proteases (**scheme 1.2**), during the acylation half of the reaction serine attacks the carbonyl of the peptide substrate, assisted by a histidine acting as a general base (**scheme 1.2(b)**), to yield a tetrahedral intermediate (**scheme 1.2(c)**). The resulting protonated imidazole (histidine side chain) is stabilized by the hydrogen bond to aspartic acid. The "oxyanion" of the tetrahedral intermediate is stabilized by interaction with the amino acid residues that surround so-called "oxyanion" hole.



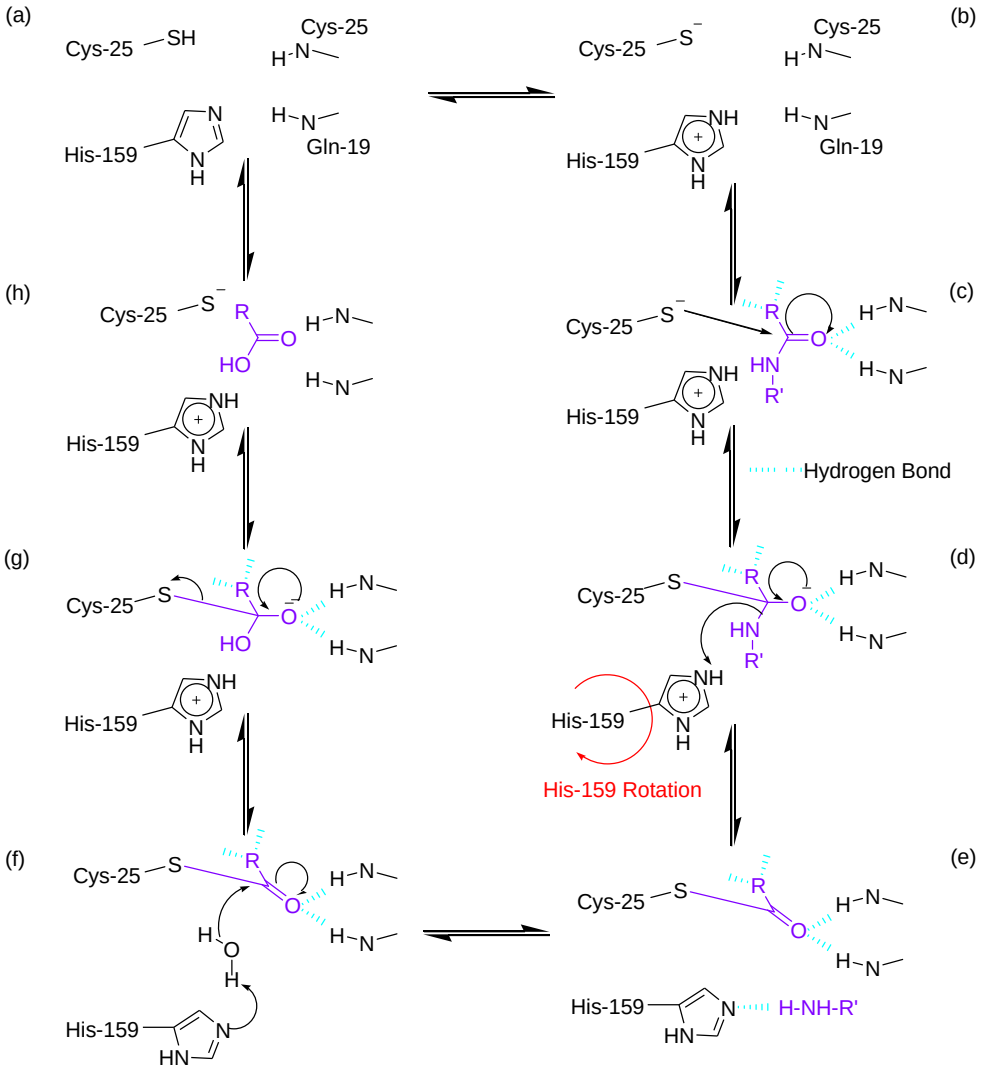
Scheme 1.2: Chymotrypsin-like catalytic mechanism.[24-26]

The tetrahedral intermediate collapses with expulsion of leaving group (**scheme 1.2(d)**), assisted by the protonated imidazole (histidine side chain) acting as a general acid, to yield the acylenzyme intermediate (**scheme 1.2(e)**). The deacylation half of

the reaction essentially repeats the above sequence: water attacks the acylenzyme (**scheme 1.2(f)**), assisted by histidine, yielding a second tetrahedral intermediate (**scheme 1.2(g)**). This intermediate collapses expelling the carboxylic acid product (**scheme 1.2(h)**). The acyl-enzyme intermediate is well established, but the tetrahedral intermediates are inferred. The transition states of the acylation and deacylation reactions resemble the high energy tetrahedral intermediates, and the terms "transition state" and "tetrahedral intermediate" are often used indiscriminately in the literature.[24-27] It is worth noting that a web of hydrogen bonding interactions links the substrate binding sites to the catalytic triad, indeed true for all proteases as the reaction proceeds, changes in bonding and charge at the scissile bond will propagate to remote enzyme sites.

1.1.3. The cysteine proteases

The cysteine proteases are all endopeptidases that differ from the serine proteases by having reactive cysteine residues. Cysteine proteases, which are mostly referred to as thiol proteases in older literature, have been found in bacteria, protozoa, plants, mammals and viruses.[28] Amongst these, is included the picorna family of viruses, a family that contains many well known human and animal pathogens, such as polioviruses (poliomyelitis), rhinoviruses (common cold), foot-and-mouth disease virus (FMDV), and hepatitis A virus.[23] Recently, cysteine proteases were also discovered in fungi.[29-31]



Scheme 1.3: Mechanism of cysteine-like proteases, using papain as an example.[32]

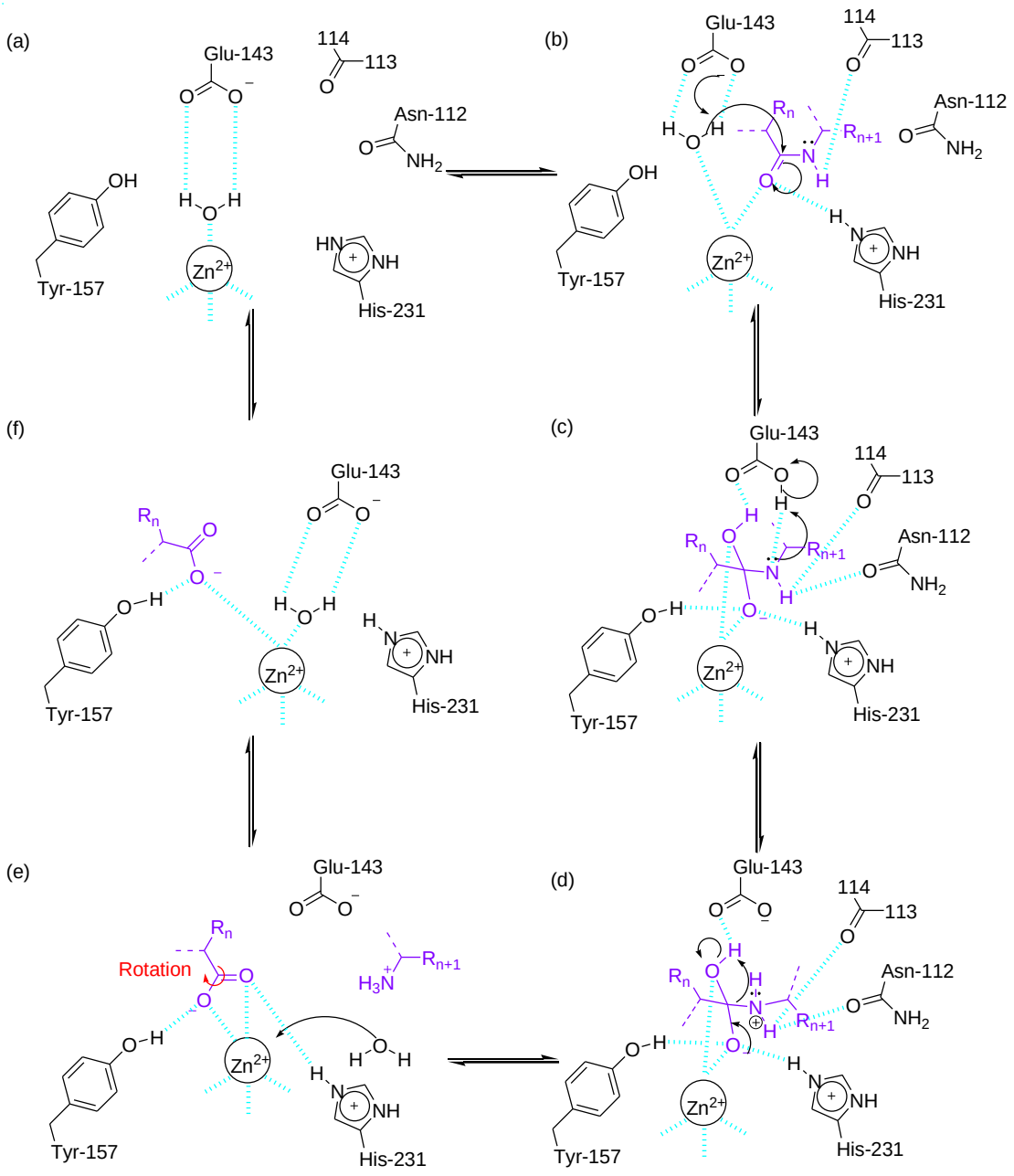
The molecular basis of the mechanism of hydrolysis catalyzed by cysteine proteases has been determined with papain. There Cys-25 and His-159 form the active site[32], but it is questionable whether these two residues are sufficient for full catalytic activity.[33] In the proposed mechanism (**scheme 1.3**), the imidazole group of the histidine polarizes the thiol group of the cysteine and enables deprotonation even at neutral to weakly acidic pH[34,35]; a thiolate/imidazolium ion pair is thereby produced (**scheme 1.3(b)**) which is highly nucleophilic. The existence of this ion pair is hardly disputed any more but whether the ion pair has actually been shown to exist is the subject of much controversy in the literature.[36,37] The ion pair mechanism does explain the unusually high reactivity of cysteine proteases toward electrophilic

reagents in comparison to the nucleophilicity of the sulfur of cysteine or glutathione, especially in slightly acidic environments.[38] The thiolate anion attacks the carbonyl carbon of the peptide bond to be cleaved (**scheme 1.3(c)**) and a tetrahedral intermediate is produced (**scheme 1.3(d)**). The oxyanion is framed by an oxyanion hole, as for serine proteases. Stabilization of this transition state occurs by hydrogen bonding to the backbone groups (**scheme 1.3(c)**). However, it was thought to be contradicted by the fact that thiono esters, in which the oxygen of the carbonyl group has been replaced by sulfur, can only be hydrolyzed by cysteine proteases and not by serine proteases. Sulfur has reduced potential to form hydrogen bonds compared to oxygen and thus the tetrahedral transition state cannot be stabilized by serine proteases. But the inability of papain to hydrolyze peptidyl thioamides clearly shows that the so-called "oxyanion hole" plays an important role in hydrolysis[39] because of the poor hydrogen bond accepting properties of sulfur. Esterification of the thiol makes the imidazolium ion sufficiently acidic (pK_a = 4, general acid catalysis) to protonate the nitrogen of the leaving group and the acyl enzyme is produced[40] (**scheme 1.3(e)**). In order to donate its proton to the amide nitrogen, rotation of the active site histidine is necessary.[32] Deacylation may occur via a general base-catalyzed mechanism, whereby the imidazole nitrogen polarizes a water molecule which then attacks the acyl enzyme at the carbonyl carbon (**scheme 1.3(f)**). The cleaved substrate (free acid) and the regenerated enzyme (**scheme 1.3(h)** to **(a)**) are produced via a second tetrahedral intermediate (**scheme 1.3(g)**). There is much discussion about whether Cys-25 and His-159 are sufficient for the hydrolysis or whether a negatively charged amino acid stabilizes the ion pair imidazolium cation/negative transition state by a symmetrical charge distribution (- + -), as for the serine proteases. For serine proteases, stabilization of the tetrahedral intermediate by a further negatively charged amino acid and an oxyanion hole are essential to favour formation of the ion pair imidazolium cation/oxyanion from an initial uncharged ground state. In cysteine proteases, an ion pair already exists and formation of the tetrahedral intermediate state only serves to redistribute charge. It follows that a cysteine protease would be more activated in the initial ground state than a serine protease or, in other words, serine proteases overcome the activation energy by

stabilization of the tetrahedral intermediate, whereas cysteine proteases start from a higher energy ground state.[41]

1.1.4. The metalloproteases

All known metalloproteases use a zinc ion (Zn^{2+}) to catalyze the hydrolysis of a peptide bond and function at neutral pH. Zinc containing proteinases are widely distributed in nature and play important roles in numerous physiological processes, such as blood-pressure regulation, as exerted by angiotensin-converting enzyme, and neprilysin, in the activation of angiotensin I and enkephalins, respectively. The matrixmetalloproteinases (MMPs) are zinc endopeptidases that are required for the degradation of extracellular matrix components, and their activities are known to be regulated by endogenous tissue inhibitors of metalloproteinases[42] (TIMPs). The unusual activation of these matrixmetalloproteinases results in diseases, such as arthritis, atherosclerosis, tumour growth, and metastasis. Due to their positive involvement in biological phenomenon, the binding modes of the inhibitors to these metalloproteinases have been sought with intentions for drug design. Other well-known zinc proteases are: collagenase; and thermolysin, a bacterial endopeptidase. Matthews and coworkers have solved several structures of thermolysin in complex with small inhibitor molecules[43], which have yielded basic insight into the cleavage mechanism[44] (**scheme 1.4**).



Scheme 1.4: Proposed thermolysin catalytic cleavage mechanism.[43,44]

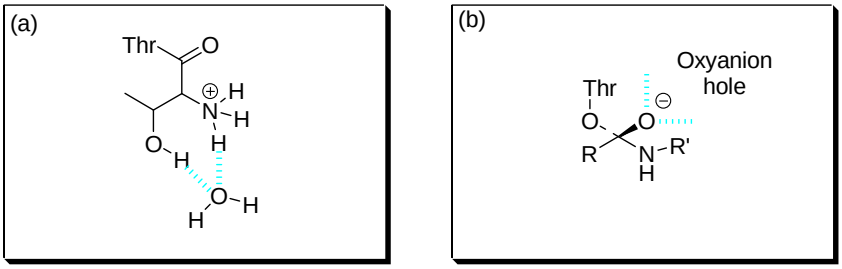
In the Michaelis complex (**scheme 1.4(b)**) the scissile peptide is tilted such that it can hydrogen bond to the zinc bound water molecule and His-231. As the scissile peptide moves toward the zinc, it may become wedged between His-231, Tyr-157, and the zinc-bound water molecule. This trajectory may cause the water molecule to move along the zinc surface toward Glu-143 and become positioned for attack on the scissile peptide. In addition to functioning as a general base, the Glu-

143 carboxyl group also acts as a proton shuttle (**scheme 1.4(c)** to **(d)**) in transferring the abstracted proton from the zinc-bound water molecule to the scissile peptide nitrogen in the tetrahedral intermediate. This dual function of the Glu-143 carboxyl group results in a "cis" addition of water across the peptide bond. Subsequently, the Glu-143 carboxyl stabilizes the tetrahedral intermediate by a salt bridge with the now positively charged nitrogen. A water molecule is then added to the scissile peptide bond in a "cis" fashion. The Tyr-157 and His-231 stabilize the tetrahedral intermediate by hydrogen-bonding. The next step of the mechanism involves cleavage of the C-N bond in the tetrahedral intermediate. The Glu-143 carboxyl group assists this step by once again functioning as a proton shuttle. This carboxyl is positioned such that it abstracts the remaining proton on the oxygen derived from the nucleophilic water molecule and delivers it to the generated primary amine. In this way, the zinc remains neutralized throughout the reaction and the amine attains the protonation state it desires at neutral pH (**scheme 1.4(d)**). This last proton transfer leaves the Glu-143 carboxyl oxygen close the oxygen of the generated carboxyl group without the ability to hydrogen bond to it. The consequent repulsion triggers a rotation of the product carboxyl (**scheme 1.4(d)**) to a monodentate orientation. This rotation would place one of the carboxyl oxygen's too close to the nitrogen of the amine product and hence "pushes" the amine product partially out of the active site, thereby reducing its binding affinity and facilitating its release. The Glu-143 carboxyl group is apparently positioned quite strategically by the enzyme since it may catalyze two proton transfers, help stabilize the transition state, and trigger the release of products. A nearby water molecule may now move into position as the fifth zinc ligand (**scheme 1.4(e)**). The resulting in pentacoordinate zinc complex may be stabilized similarly to the tetrahedral intermediate. Departure of the carboxyl containing cleavage product and movement of the zinc-bound water molecule to the native position completes the catalytic cycle (**scheme 1.4(f)** to **(a)**).

1.1.5. The threonine proteases

Threonine proteases have many common active site features with serine and cysteine proteases, including an active site nucleophile and a general base, which are

often the target of irreversible inhibitors. One of the best known threonine proteases is the 20S proteasome.[45] The 20S proteasome is a 6500 amino acid protein with an active site N-terminal threonine, and it plays a central role in protein degradation in eukaryotic cells. The active site residues of threonine proteases have many mechanistic features in common with serine and cysteine proteases. Each enzyme has an active site nucleophile and a basic residue, which can also function as a general acid in the catalytic mechanism. The transition states for serine, cysteine, and threonine proteases all involve formation of a tetrahedral intermediate. The oxyanion of the tetrahedral intermediate is frequently stabilized by interaction with several hydrogen bond donors, which is commonly referred to as the oxyanion hole[46] (**scheme 1.5**).



Scheme 1.5: (a) Active site residue of threonine in the 20S proteasome; (b). transition state for hydrolysis.[46]

1.1.6. The glutamic acid proteases

The glutamic protease family has recently been re-classified as a sixth catalytic type of peptidase. Previously known as the A4 family of aspartic endopeptidases, recent analysis of the molecular structure and catalytic mechanism has identified these enzymes as a novel protease family. The Eqolisins, a name derived from the active-site residues, glutamic acid (E) and glutamine (Q)[47], have a previously undescribed ß-sandwich tertiary fold and a unique catalytic dyad consisting of glutamine and glutamate residues which, respectively, activate the nucleophilic water and stabilise the tetrahedral intermediate on the hydrolytic pathway.[47]

1.2. Protease substrate specificity

Proteolytic enzymes have many roles, for example they are important in controlling the cleavage of prohormones generating active hormones which regulate many physiological processes.[6] Proteases play an important role in the control of blood coagulation, whose malfunctioning can have dramatic consequences. In the case of apoptosis, caspases inactivate proteins responsible of cell maintenance and DNA repair, therefore leading to the death of the cell. Many other examples could be given as proteases are essential to many key processes of living organisms and are essential to cellular life and activity.

From a therapeutic point of view, proteases account for 5-10% of all drug targets, with applications in cancer, blood coagulation, diabetes, hypertension, HIV and other viral processes and neurodegenerative diseases.[6] They therefore represent a series of interesting targets, and much effort has concentrated in investigating protease substrate specificity. This knowledge helps us to get a better understanding of the mode of action of proteases and rapidly leads to inhibitors.

A characteristic function of proteases is their ability to discriminate among many potential substrates, their-so-called substrate specificity, which is a critical factor in maintaining the fidelity of the biological processes in which a protease acts. For researchers, substrate specificity can serve as a handle by which proteases can be discriminated from others in its class, even in cases in which a large degree of structural homology exists. Moreover, the knowledge of the substrate specificity of a protease can greatly enhance the understanding of its function, since it helps in placing the enzyme in its physiological environment. The protease's natural substrate can therefore be identified and the action of the enzyme can be more easily included in a cascade of reactions. From another point of view, the discovery of new selective substrates of the protease can also be very interesting since it allows the elaboration of protease inhibitors by the use of peptidomimetic strategies to build up the candidate.

The investigation of the substrate specificity of a protease consists of the assessment of its preference regarding the P and P' amino acid positions of a peptide chain. This allows establishment either of a unique substrate for very specific

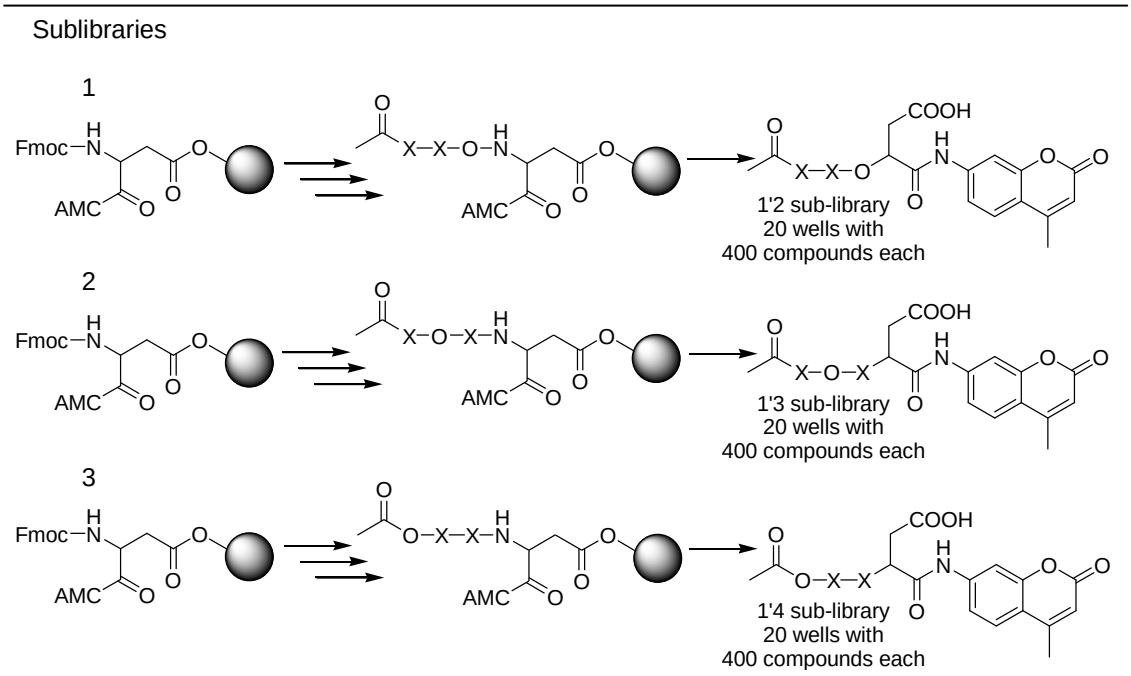
proteases, or trends regarding the affinity of amino acids for the S and S' subpockets (hydrophylicity, size allowance, etc.) for more general proteases. Several methods have been used so far to analyse protease substrate specificity, and the following techniques have been reported.

1.3. High-Throughput Proteases Screening Methods

1.3.1. Positional-scanning assays

Positional-scanning synthetic combinatorial libraries[48] (PS-SCL) of fluorogenic peptide substrates are powerful tools for determining protease specificity. In contrast to other combinatorial libraries, this library format provides rapid and continuous information on each of the varied substituents in the substrate. Peptide-based positional-scanning synthetic combinatorial libraries are generally composed of several sub-libraries in which one position is defined with an amino acid, and the remaining positions contain a mixture of amino acids present in approximately equimolar concentrations and the analysis of each sub-library will identify the preferred amino acids for each position (i.e. …, P_2, P_1, P_1', P_2', …).

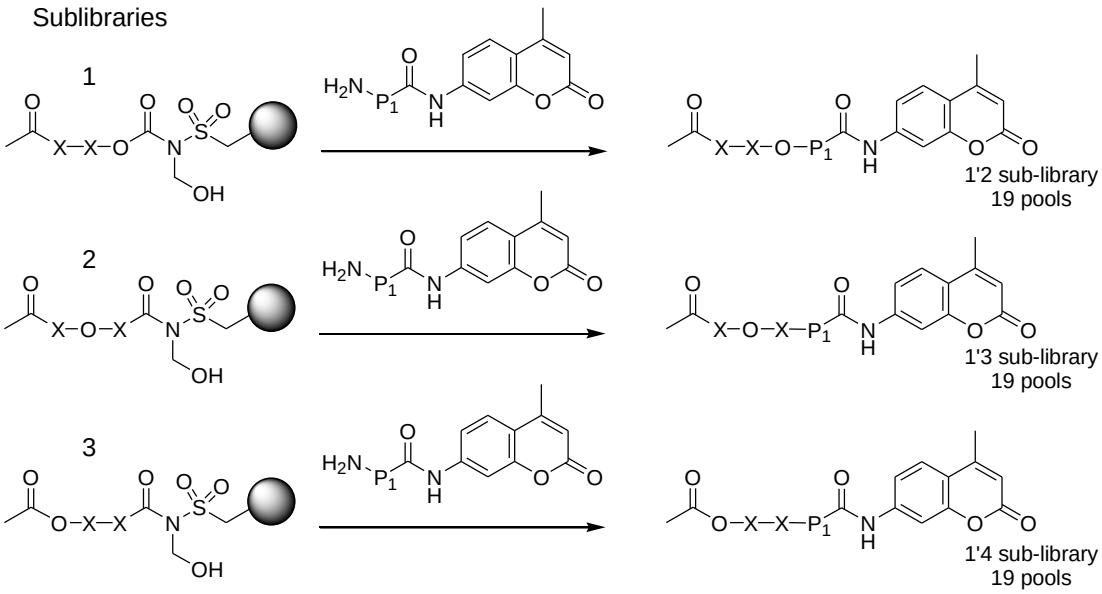
A positional-scanning library was prepared by Rano *et al.*[49] with the general structure Ac-X_1-X_2-X_3-Asp-AMC (AMC = 7-amino-4-methylcoumarin) to rapidly and accurately assess the P_4-P_2 specificity for interleukin-1β converting enzyme (ICE) that requires aspartic acid (Asp) in the P_1 position.[49,50] Specific cleavage of the amide bond after the Asp residue liberates a fluorescent AMC leaving group, thus allowing for the simple determination of cleavage rates for a library of substrates. The P_1 Asp-coumarin substrate was conveniently linked to an insoluble polymer through the Asp carboxylic acid side chain, which allowed for library synthesis by standard peptide synthesis techniques (**scheme 1.6**). The particular PS-SCL used consisted of three separate sub-libraries that were each composed of 8000 compounds.



Scheme 1.6: Tetrapeptide-AMC combinatorial positional-scanning library. The positional-scanning library consists of 3 separate sub-libraries of 8000 compounds (20 wells of 400 compounds each). Each library is comprised of an amino-terminal acylated tetrapeptide –aminomethylcoumarin derivative, where "X" represents the mixture of proteogenic amino acids and "O" represents the defined amino acid (excluding cysteine and methionine, which were replaced by norleucine and D-alanine).[49]

This library was designed on the basis of several catalytic properties of the enzyme IC. Firstly, its most distinctive feature is a near absolute requirement for aspartic acid in the position to the immediate left of the scissible bond (P_1) of both peptide and macromolecular substrates. Secondly, there is an equally stringent requirement for four amino acids on the amino-terminal side of the scissile bond. Finally, tetrapeptides terminating in -Asp-AMC are highly sensitive fluorogenic substrates for the enzyme, being specifically cleaved after Asp to liberate the fluorescent leaving group, AMC. Using this strategy, analysis of the three sub-libraries showed that histidine at P_2, glutamine at P_3 and tryptophan at P_4 as preferred amino acids across the S_2, S_3 and S_4 subsites.[49] This study, the first one to report using PS-SCL to determine protease specificity, provided the evidence that developing a general strategy to incorporate all 20 proteinogenic amino acids at every P position of a positional-scanning synthetic combinatorial library would allow the specificity of virtually any protease to be rapidly determined. With this aim in

mind, Ellman *et al.*[51] developed a positional-scanning synthetic combinatorial substrate library to define the substrate specificity of plasmin and thrombin. The limitation of Rano's[49] method which required linkage to the solid support through the side chain of the P_1 substituent (i.e. aspartic acid) was eliminated in Elmann's[51] method by using a condensed fluorogenic AMC P_1-amino acid derivative which could be added through nucleophilic addition to support-bound sublibraries, cleaving the sulphonamide linker, releasing the sublibraries, allowing for complete randomization at the P_1 position (**scheme 1.7**).



Scheme 1.7: Preparation of three sublibraries (X = all natural amino acids; O = defined amino acid). Each made up of 19 wells containing 361 compounds by segment condensation reaction with a P_1 fluorogenic amino acid substrate.[51]

The resulting enzymatic profiles from the library resembled the known physiological cleavage sites of these enzymes and were verified by single-substrate kinetic analysis (**figure 1.4**). Potential substrate recognition determinants on the enzymes were then identified through the three-dimensional modelling of the substrates in the active sites of the enzymes.

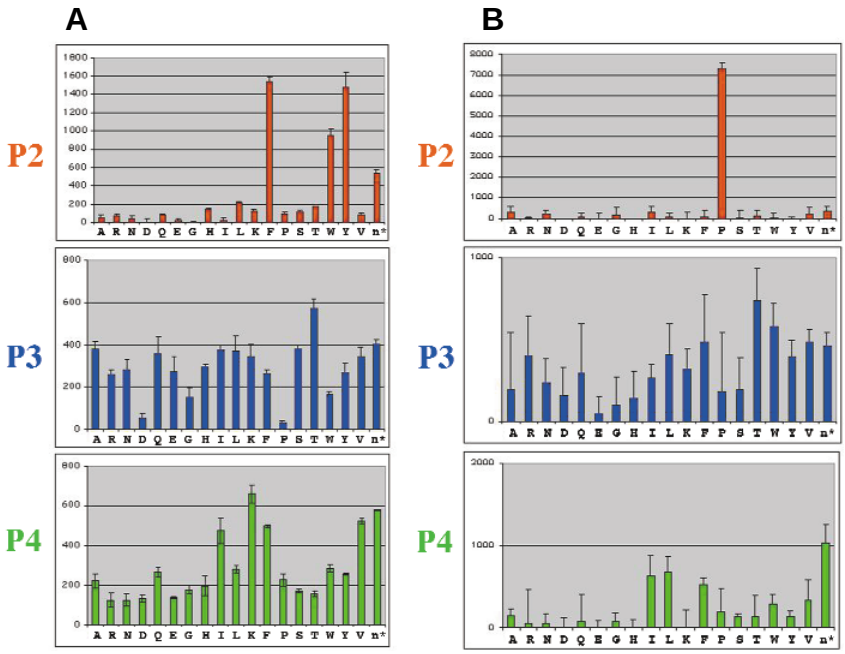


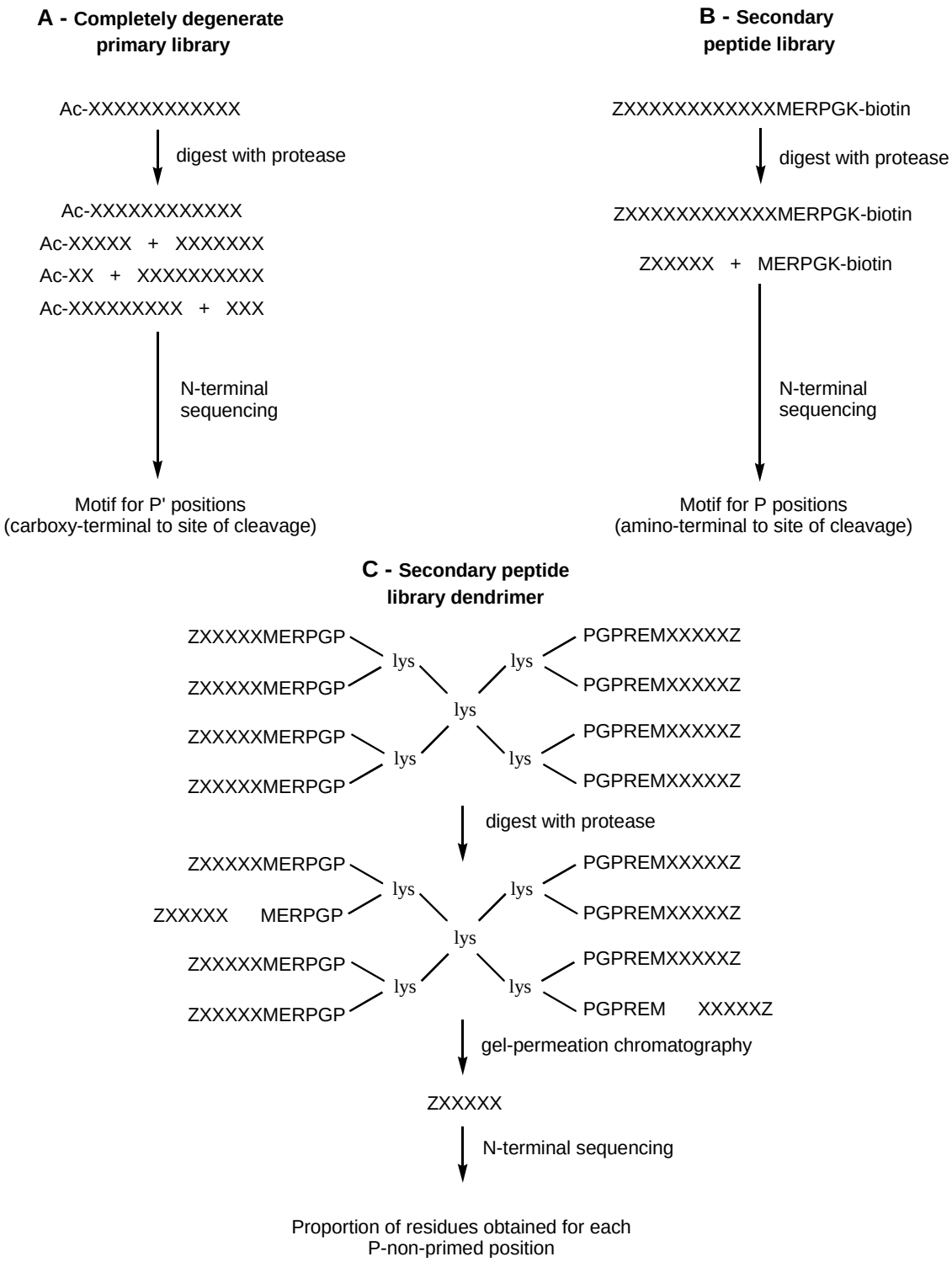
Figure 1.4: (A) Activity of plasmin in a P_1-Lys positional-scanning synthetic combinatorial library. (B) Activity of thrombin in a P_1-Lys positional-scanning synthetic combinatorial library. y-axis indicates pmoles of fluorophore released per second. x-axis indicates the amino acid held constant at each position, designated by the one-letter code (n* represents norleucine) (reproduced with permission).[51]

Results from this study show that, in contrary to the traditional assumption that plasmin possesses broad substrate specificity, this protease demonstrates a distinct preference for aromatic amino acids at P_2 and a moderate preference for lysine and hydrophobic amino acids at P_4. Preference for both charged and aromatic amino acids is also known for the S_2 pocket of the cysteine protease cruzain, where a glutamate side chain adjusts to accommodate the disparate side chains (**table 1.1**).

Table 1.1: Cleavage sites in the physiological substrates of plasmin.[51]

	P_4	P_3	P_2	P_1
Optimal Plasmin Substrate Specificity from PS-SCl	K/V/I/F	X	F/Y/W	R/K
Vitronectin	K	G	Y	R
Osteocalcin	E	A	Y	R
Factor X	I	T	F	R
PAR1 (70)	T	E	Y	R

The development of tools permitting the preparation of pools of synthetic peptides has made possible the screening of combinatorial mixtures of non-fluorogenic N-acetylated peptides by incubation with proteases.[52-54] In these approaches, completely degenerate peptide mixtures were digested with the protease of interest. After reaction, the substrates that were cleaved by the enzyme were left with a free amino group and could therefore be submitted to Edman degradation[55] (technique that allows sequence determination for any peptide with an unprotected α-amino group). This provides information about the prime side specificity of the enzyme, the first round of sequencing giving information on P_1', the second on P_2', etc (**scheme 1.8**). Selectivity data derived from this initial library screen was then used to design additional peptide libraries that could be used to determine the cleavage motif for the non-primed positions. These secondary libraries were prepared in which a number of degenerate positions were followed by positions in which the optimal primed residues were fixed, to direct cleavage to a single bond within the peptide sequence. Secondary libraries were made with a biotin group at the carboxy-terminus and with a free (unblocked) amino-terminus. Following library digestion with the enzyme of interest, the unreacted peptides and carboxy-terminal fragments could be removed with immobilized avidin. The amino-terminal fragments remaining in solution can then be sequenced by Edman degradation[55] allowing the determination of non-primed preferences.[52,53] Reiteration of this process allows an optimal recognition sequence to be determined. Later, a more elaborate second step in this methodology was introduced using a degenerate peptide dendrimer for determination of specificity at non-primed positions, and the complete cleavage specificity for μ-calpain was determined for all positions flanking the cleaved peptide.[54]



Scheme 1.8: Schematic representation for the determination of specificity at primed (A) and non-primed positions (B and C).[52,54]

In a more recent approach, combinatorial fluorogenic and standard peptide libraries combined with fluorescence and LC/MS/MS-based assays were used to analyze the substrate specificity of the lysosomal protease tripeptidyl-peptidase[56] (TPP I). The relative contribution of all non-sulfur-containing amino acids and norleucine was determined for each of the P_3, P_2, P_1, P_1', and P_2' substituted residues. Over 7,200 different fluorogenic and standard peptides (as individual peptides or within mixtures) were synthesized for the systematic analysis of the P_3, P_2, P_1, P_1', and P_2' residues. Reactions were conducted in flat-bottom polystyrene 96-well plates and initiated by adding 80 μl of peptide solution to 20 μl of TPP I solution and the digestion rate was determined by measuring the initial rate of digestion (using conditions where the rate of digestion was linear) with respect to time and enzyme concentration. The P_2-diverse library was restricted at the P_1 position to the following four optimum residues found when evaluating a starting P_1 library I: leucine, norleucine, phenylalanine, and tyrosine. Here, for each of these fixed P1 residues, 19 pools of 19 ACC tripeptides were synthesized, each with a different fixed P_2 position and containing a randomized mixture of 19 residues at the P_3 position. These 76 peptide pools were evaluated by using the fluorescence assay to determine the relative contribution of each P_2 residue in the context of four different P_1 residues averaged over all P_3 residues. Subsequently, after analysis of the P_3 library and validation of an LC/MS/MS-based assay, a subset of the peptide pools to determine the specificity constants for 589 individual peptides from 31 pools in the P_2 library was analysed, allowing more thorough evaluation of potential P_3-P_2-P_1 interactions. The P_1'P_2' pentapeptide library was digested and the TPP I concentration during incubation was 2.1 nM with three ACC peptides (Ala-Ala-Phe-ACC, Asn-Nle-Nle-ACC, and Ala-Ala-Pro-ACC) included as internal standards with a final concentration of 1 μM. Samples (50 μl) were removed at 4, 40, and 480 min, and the reactions were terminated by adding 5 μl of 15% trifluoroacetic acid. Terminated reactions were diluted 5-fold with distilled water before LC/MS/MS analysis. Before the digestion of the P_1' tetrapeptide library, equimolar amounts of 19 P_1' tetrapeptides were mixed. The peptide mixture was digested under the same conditions as the P_1'P_2' pentapeptide library, but 7 time points were monitored (20 s and 4, 10, 40, 120, 280, and 480 min) to measure the time course of digestion. To

summarize this large amount of data, a substrate specificity matrix that compares relative substrate specificities of different residues at each position was constructed, with the optimal residue being assigned a score of 1 (**table 1.2**).

Table 1.2: Substrate specificity for each amino acid analysed using the LC/MS/MS assay and normalised to the corresponding best amino acid residue (reproduced with permission).[56]

	P_3		P_2		P_1		P_1'		P_2'
Arg	1.00	Nle	1.00	Leu	1.00	Phe	1.00	Val	1.00
His	8.63E-01	Pro	5.19E-01	Phe	9.01E-01	Trp	9.01E-01	Ile	9.68E-01
Tyr	4.74E-01	Ala	5.10E-01	Nle	8.85E-01	Tyr	8.06E-01	Nle	9.43E-01
Lys	3.51E-01	Phe	2.89E-01	Tyr	2.14E-01	Val	7.33E-01	Leu	9.20E-01
Gln	2.70E-01	Tyr	2.72E-01	Trp	1.04E-02	Ile	6.07E-01	Trp	7.20E-01
Phe	1.71E-01	Trp	1.94E-01	Glu	4.46E-02	Ala	2.91E-01	Phe	6.90E-01
Ile	1.36E-01	Ile	1.80E-01	Asp	2.31E-02	Thr	2.56E-01	Tyr	6.57E-01
Ala	1.14E-01	Val	1.79E-01	Gln	1.91E-02	Leu	1.94E-01	Ser	6.11E-01
Ser	1.08E-02	Thr	1.33E-01	Ala	1.52E-03	Nle	1.91E-01	Thr	6.04E-01
Glu	9.12E-02	Glu	1.32E-01	Asn	5.54E-03	Ser	1.34E-01	Ala	5.63E-01
Thr	9.10E-02	Leu	1.00E-01	Arg	3.15E-03	Glu	1.05E-01	Arg	5.26E-01
Val	8.60E-02	Ser	9.55E-02	Lys	2.79E-03	Asn	1.04E-01	Gln	5.17E-01
Asn	8.09E-02	Gly	6.78E-02	Ser	2.30E-03	Asp	8.92E-02	His	4.25E-01
Nle	6.67E-02	Gln	6.48E-02	His	2.00E-03	His	6.94E-02	Asn	4.03E-01
Trp	5.84E-02	Lys	2.55E-02	Val	1.48E-03	Arg	5.32E-02	Lys	3.46E-01
Leu	5.31E-02	Asn	2.04E-02	Ile	6.97E-04	Gln	4.90E-02	Glu	2.88E-01
Gly	4.29E-02	Asp	1.52E-02	Thr	6.47E-04	Gly	2.31E-02	Asp	2.85E-01
Asp	3.06E-02	Arg	9.40E-03	Gly	1.02E-04	Lys	2.26E-03	Gly	2.64E-01
Pro	4.87E-03	His	8.56E-03	Pro	1.35E-05	Pro	1.42E-03	Pro	1.10E-01

A plot of the relative k_{cat}/K_M values indicates that the P_1 position is the most critical, followed by P_1', P_2, and P_3, with the P_2' position tolerating a wide range of substitutions other than Pro (**figure 1.5**). Considering each position independently, assignment of a relative score to all possible pentapeptide sequences is made as the product of the relative substrate specificities at each position (**figure 1.5**).

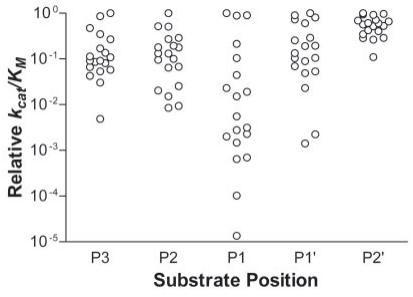


FIGURE 1.5: Distribution of relative substrate specificities at different positions. The plot shows the log value for each residue calculated as described in Table 1.2 (reproduced with permission).[56]

However, the analyses of the P_2 and P_2' libraries and P_1 library II revealed that there are peptides that exhibit synergistic or deleterious interactions between residues at different positions. For instance, the beneficial effects that were found for positively charged P_3 residues may be related to the fact that these residues are located at the N terminus of the peptides. For P_3 residues, the positive charge of the N terminus, coupled with the positive charge of the side chain, may enhance binding to a negatively charged subsite on TPP I. Thus, a more sophisticated approach, including consideration of structural and electrostatic properties of both substrate and enzyme, may be required to predict TPP I substrate specificity. Nevertheless, this study has enabled the synthesis of new fluorogenic peptides that exhibit greater selectivity for TPP I compared with substrates reported previously.[56]

1.3.2. Bead assays

The first account of substrate conversion by proteases on solid supports monitored in a library format was the use of FRET-quenched substrates, which upon interaction with proteolytic enzymes cleaved to release a quencher and left a fluorophore remaining on the solid support[57] (**figure 1.6**). The enzyme activity towards a particular substrate was essentially proportional to the amount of fluorescence observed on the support when a complete split–mix library of millions of fluorescence quenched substrates was subjected to enzyme hydrolysis.[58] In assays of enzymes the choice of resin polymer material is crucial for the success of the assay. Since the enzyme must be able to penetrate the polymer network and move freely within the interior of the bead, the polymer itself should not have any significant interaction with the protein. The porosity of the resin should therefore match the size of the enzyme investigated. Initially Kieselguhr resin[59] was used, but later PEG-based resins[60], such as PEGA 1900 and PEGA 4000, which may be used for enzymes with a molecular weight of up to 60,000 and 120,000, respectively, have become popular. This solid-phase split-mix approach with FRET substrates has been adapted by a number of groups. The screening of 500,000 FRET substrates on PEGA 1900 with the parasitic cysteine protease cruzain yielded substrates with a high specificity for basic residues in P_5, P_3, P_1 and P'_2 or P'_3.[61]

Rossé *et al.*[62], using the same approach, developed the Dabcyl-Lucifer yellow FRET pair to produce substrate libraries on PEGA1900 resin for the analysis of the substrate specificity of *E. Coli* leader peptidase and for napsin A. In this assay the hydrophilic character of the sulphonated Lucifer yellow helped solvation of the substrates and the longer wavelengths of excitation (488 nm) and emission (520 nm) brought the assay well above the spectral region of interference from smaller chromophores in the substrate. The method was used to determine the substrate specificity of the scarce napsin A in crude cell extracts. It prompted the development of "one-bead-two compounds" combinatorial assays in which each bead could be considered a separate assay container for binding versus chemical transformation. In these novel kinds of assays a secondary compound that is uniformly linked to all beads in the library is used to monitor the reaction or interaction with a library member on each bead. While the primary compound is the inhibitor the assays are immediately useful for the detection of new inhibitors from large combinatorial libraries using a large variety of enzymes.[63] In each bead of the library the enzyme competes for binding to the substrates and a putative inhibitory library component bound to the bead. If the compound inhibits cleavage of the substrate, which most conveniently should be a FRET substrate with a high affinity for the binding site, the substrate remains intact and no fluorescence is observed. On the other hand, with beads with no inhibitors the substrate is rapidly cleaved resulting in removal of quencher and development of strong fluorescence from the resin-bound fluorophore. The use of FRET substrates has the additional advantage that the hits may be isolated using automated bead sorting. Release of the inhibitor on isolated hits and structure elucidation either by a variety of encoding–decoding techniques (eg. mass spectrometry analysis) allowing the rapid and efficient identification of novel enzyme inhibitors.

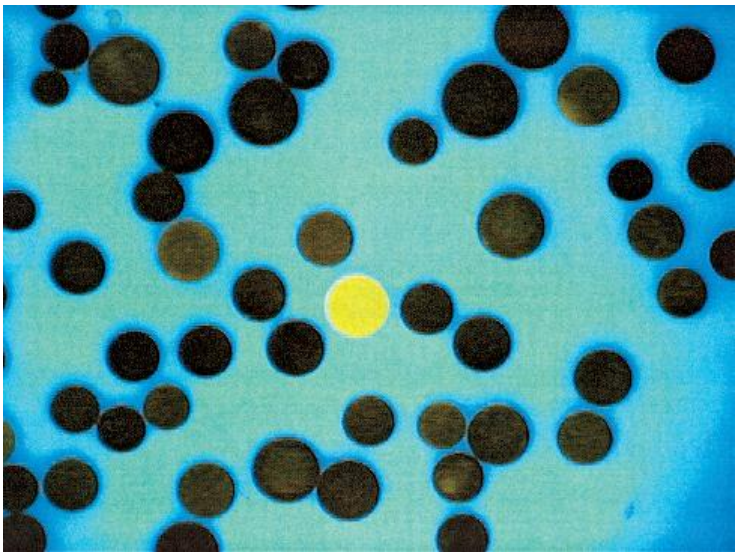


Figure 1.6: Image of beads from the peptide library after incubation with *E. coli* leader peptidase. Small portions of the beads were placed on a glass plate and examined under a fluorescence microscope (reproduced with permission).[63]

1.3.3. Microarray assays

The use of internal fluorescence quenching is a very efficient method for determining protease substrate specificity. The basis of the approach is the preparation of a peptide labelled on one end with a fluorophore donor and on the other end with an acceptor quenching the fluorescence. Upon cleavage of the peptide, a significant increase in fluorescence is observed. The combination of FRET detection with PNA encoding of the different peptidic sequences on a microarray slide as carried out by Diaz-Mochon *et al.*[64], has significantly increased its power. The encoding of each amino acid by a PNA triplet allowed 10,000 peptides arising from a split and mix solid phase synthesis to be screened against chymopapain and subtilisin, using FAM/TAMRA as the FRET pair (**figure 1.7**). The use of PNA codes allowed enzymatic assays to be carried out on the cleaved labelled peptides in solution, followed by hybridisation on a DNA microchip that permitted the straightforward identification of each sequence. Moreover, the use of the microarray format allowed minimal consumption of enzyme (60 pmole) and library (3.5 nmole).

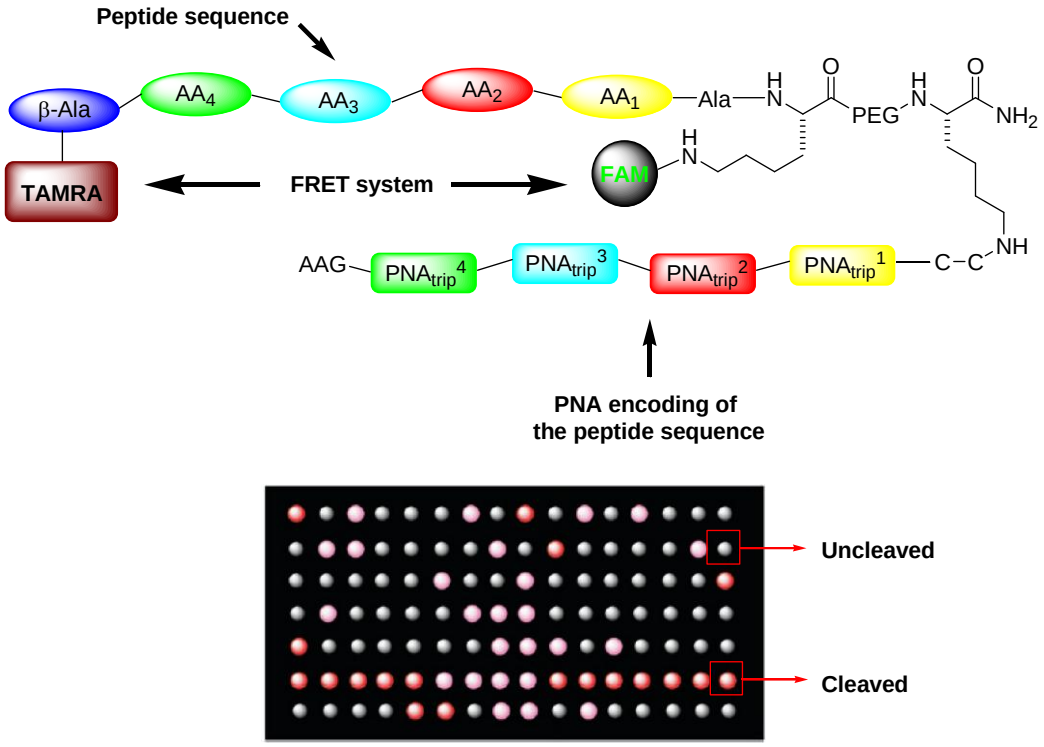


Figure 1.7: General structure of the PNA encoded FRET based peptide libraries to measure protease activity in solution followed by hybridization to DNA array to provide fluorescent spots where the cleaved substrates were bound.[64]

Winssinger and co-workers also used DNA micro-arrays to monitor protease function by capture of hybrid molecules composed of PNA and a protease inhibitor or substrate.[65-68] In their functional assays a fluorescent molecule was linked to a library of putative active small molecules via a strand of PNA matching a position on the DNA array and encoding for the small molecule structure. After binding to a target in a biological sample and removal of unbound material by dialysis, hybridisation of remaining protein bound ligand-PNA constructs to the array provided active ligands as identified from the positions of the fluorescent spots on the array (**figure 1.8**). The technique was further developed for substrate specificity determination using a split mix approach to construct substrate libraries linked to the encoding PNA via a rhodamine probe. The substrates were attached to the fluorophore via amide bonds and upon hydrolysis of this bond the fluorophore increased its fluorescence up to ≈1000-fold in intensity thus providing quantitative information on the *N*-terminal substrate preferences upon hybridisation of the mixture to the DNA array. This technique was successfully used to characterise the

differences in thrombin, plasmin and caspase 3 specificity[66], as well as in the profiling of dustmite allergens.[65]

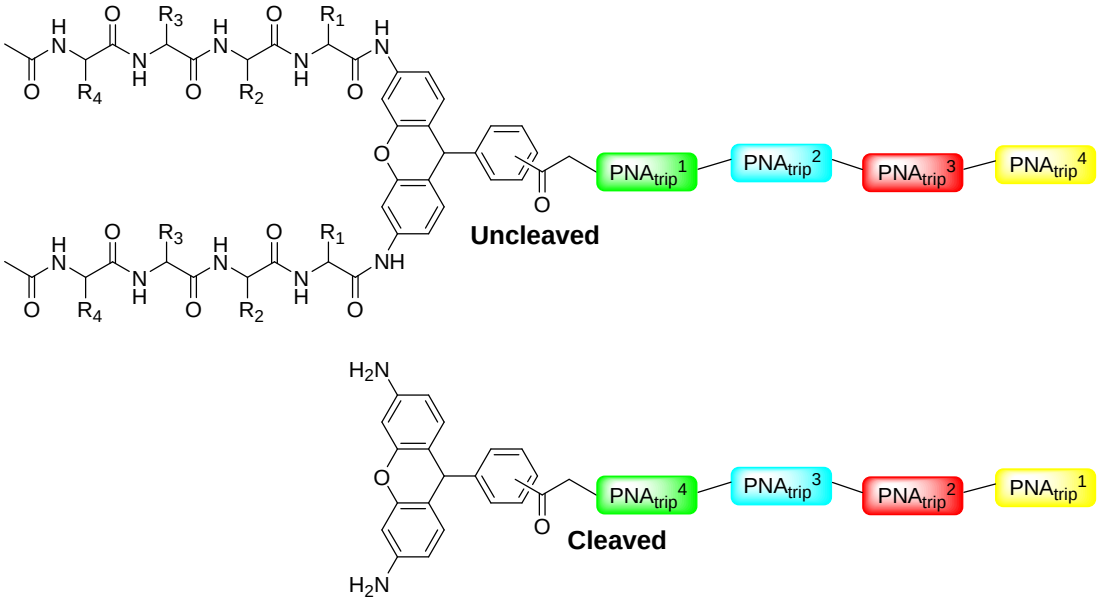


Figure 1.8: Peptide substrates conjugated to PNA via rhodamine amide allowed Winsinger's group to measure protease activity in solution followed by hybridization to DNA array to provide fluorescent spots where the free rhodamine was bound.[65-68]

The elaboration by Ellman *et al.*[69] of the bifunctional fluorophore 7-amino-methylcoumarin (AMC) permitted the direct attachment of the fluorophore linked peptide to a solid support. The method allowed variations to be made at the P_1 position (**figure 1.9**) with the 7-amino-4-carbamoylmethylcoumarin (ACC) fluorophore showing kinetic profiles comparable to AMC but with up to a three fold fluorescence increase. Despite harsh conditions being required to achieve completion of the first coupling to the aniline because of to its poor nucleophilicity[70], applications of these fluorogenic substrates have been successfully made to profile substrate specificity of the nonprime side for multiple proteases, such as serine proteases. In one approach to fluorogenic array based substrate specificity profiling, Ellman's group[71] reported the use of ACC peptides with proteolytic assays directly carried out on the 2D surface. The bifunctional character of the ACC fluorophore indeed permitted its attachment to the slide by means of a specific linker and oxime formation with aldehydes functionalities of the glass slide.[71] Substrate specificity was obtained and compared well to solution phase assays for a variety of proteases.

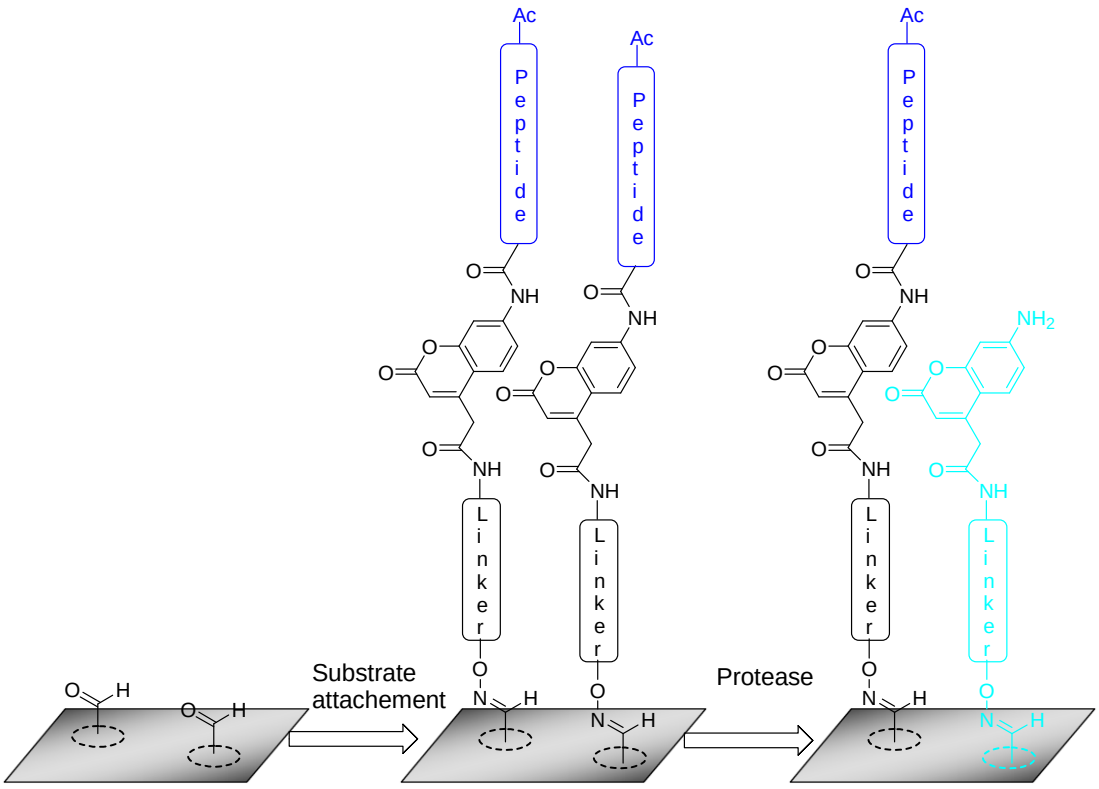


Figure 1.9: Fluorogenic AMC peptide cleavage.[69]

A library of substrates was synthesized, printed, and assayed to demonstrate the utility of peptide-ACC microarrays for the rapid determination of protease substrate specificity. The 361-member spatially separate library contained a fixed P_1-Lys and P_4-Ala with all combinations of proteinogenic amino acids (except cysteine) at the P_2 and P_3 sites. To serve as standards for quantitation of fluorescence, multiple copies of unacylated ACC and acetyl-capped ACC were also printed on the array surface. The array was then used to determine the extended substrate specificity of the serine protease thrombin by adding 15 μL of a 250 nM solution of the enzyme to the array surface followed by incubation for 60 min. To validate the results, four of the peptidyl coumarin substrates were synthesized, purified, and assayed in solution (**table 1.3**). As the data shows, the k_{cat}/K_m values for the substrates on the array are comparable to solution-phase results. Of note is the fact that considerably less enzyme is needed to determine the specificity profile on the array relative to other methods, due to the miniaturized nature of the array and the increased sensitivity of

fluorescence detection that results from the high substrate density on the array surface.

Table 1.3: Comparison of the Relative k_{cat}/K_m for purified underivatized ACC-peptides in solution and the corresponding library substrates on the array, upon treatment with thrombin (data have been normalized to the Ac-ATPK-ACC value).[69]

substrate	solution phase k_{cat}/K_m	microarray k_{cat}/K_m
Ac-ATPK-ACC	1.00	1.00
Ac-AGPK-ACC	0.27	0.26
Ac-ADAK-ACC	0.02	0.09
Ac-AFSK-ACC	0.02	0.00

An interesting approach to protease analysis was the introduction of fluid-phase microarrays by Diamond *et al.*.[72] In this methodology, chemical compounds within individual nanolitre droplets of glycerol were microarrayed onto glass slides. Using subsequent aerosol deposition, reagents and water were deposited into each reaction centre to rapidly assemble diverse multicomponent reactions without cross-contamination or the need for surface linkage. This technique allows the kinetic profiling of protease mixtures, study of protease-substrate interactions, and high-throughput screening. To demonstrate the feasibility of a nanolitre-scale screening assay of serine proteases, a 16×16 microarray of glycerol was generated with the central positions (boxed in red, **figure 1.9(A)**) spiked with increasing doses from 0 to 100 mM of benzamidine, a thrombin inhibitor. Human thrombin at 10 units/ml (400 nl/s for 4 s) was deposited on the microarray followed by deposition of boc-VPR-MCA (MCA - methylcoumarin-7-amide) at 10 mM in DMSO at 400 nl/s for 4 s. The microarray was incubated and imaged (**figure 1.9(A)**) to produce a dose-response curve with an IC50 of ≈5 mM (**figure 1.9(B)**) for benzamidine on thrombin in a high glycerol background. An IC50 of 5 mM was found using similar reaction conditions in glycerol in a 96-well plate assay.

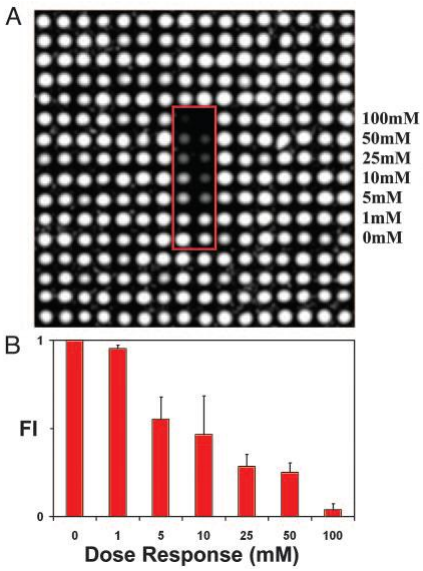
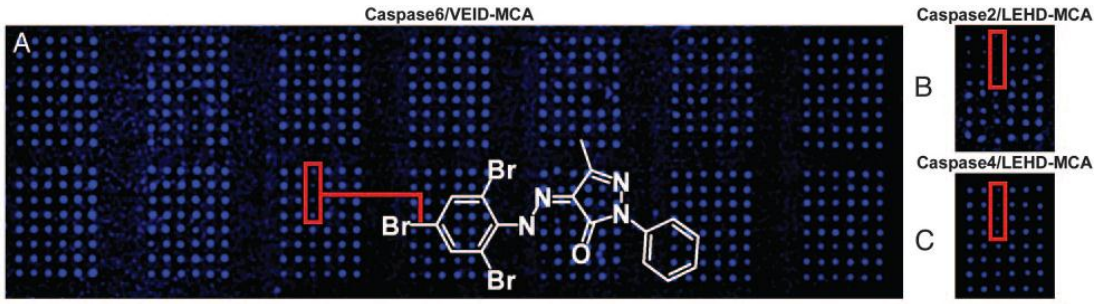


Figure 1.9: (A) Microarrays activated with human thrombin and boc-VPR-MCA displayed a dose-dependent inhibition by the thrombin inhibitor benzamidine (B) As seen in the boxed central zone that contained duplicate reaction centers with increasing concentration of the inhibitor from 0 to 100 mM. All other reaction centers lacked inhibitor allowing full fluorescence generation (reproduced with permission).[72]

A commercially available exploratory library of 352 diverse compounds supplied in 2-µl samples at 15 nmol/µl (15 mM in DMSO) was microarrayed consuming ≈1 nanomoles of each compound. Replicate slides with each compound were prepared and arrayed at 1 mM in glycerol in quadruplicate along with 32 blanks to serve as positive controls for uninhibited reactions. A microarray was sprayed sequentially with human caspase 6 and then its substrate VEID-MCA. The final concentrations in the glycerol screening reactions were 0.227 units/µl caspase and 11.36 µM peptide substrate, based on the delivery efficiency for the spraying of each reagent (400 nl/s) into each spot. Four spots containing the same compound on the microarray displayed low substrate conversion (**figure 1.10(A)**). This compound, 5-methyl-2-phenyl-4-[(2,4,6-tribromo-phenyl)-hydrazono]-2,4-dihydropyrazol-3-one (compound **1.1**, 515.0 *M*r), also reduced substrate conversion when microarrays were screened with human caspase 2 and LEHD-MCA (**figure 1.10(B)**) or human caspase 4 and LEHD-MCA.

Inhibitors of caspases 2, 4, and 6 were identified using a 352-compound library microarrayed in quadripicate on 100 slides and screened against the caspases as well as thrombin and chymotrypsin (**figure 1.10**).



1.1

Figure 1.10: (A) Nanolitre screening of a compound library microarrayed at 1 mM in quadruplicate with human caspase 6, and a fluorogenic substrate (VEID-MCA) identified the location of an inhibitor (compound 1.1). When replicate microarrays were screened against caspase 2 (B) or 4 (C), the identical location on the microarrays as seen in (A) displayed low-fluorescence emission indicative of enzyme inhibition. When tested in triplicate in a standard well plate assay, compound 1.1 caused a dose-dependent inhibition of caspases 2, 4, and 6 with an IC50 of ≈0.5–5 mM against the three caspases (reproduced with permission).[72]

Later, the same methodology was applied to a 722-member library of fluorogenic protease substrates. The library, of the general format Ac-Ala-X-X-(Arg/Lys)-ACC was synthesized (X - all natural amino acids except cysteine) and microarrayed with fluorescent calibration standards (unacylated ACC, acetyl-capped ACC, and blanks) in glycerol nanodroplets on glass slides. Specificities of 13 serine proteases (activated protein C, plasma kallikrein, factor VIIa, factor IXaβ, factor XIa and factor α XIIa, activated complement C1s, C1r, and D, tryptase, trypsin, subtilisin Carlsberg, and cathepsin G) and 11 papain-like cysteine proteases (cathepsin B, H, K, L, S, and V, rhodesain, papain, chymopapain, ficin, and stem bromelain) were obtained from 103,968 separate microarray fluorogenic reactions (722 substrates × 24 different proteases × 6 replicates). This was the first comprehensive study to report the substrate specificity of rhodesain, a papain-like cysteine protease expressed by *Trypanasoma brucei rhodesiense*, a parasitic protozoa responsible for causing sleeping sickness. Rhodesain displayed a strong P_2 preference for Leu, Val, Phe, and Tyr in both the P_1 = Lys and Arg libraries.[73,74]

A different approach was developed using fluorogenic peptides trapped in a 3D hydrogel.[75] This chip consisted of a supramolecular structure made from a low-molecular weight hydrogelator based on a glycosylated amino-acid scaffold. The supramolecular gel, formed at low critical concentration of the hydrogelator, lead to a semi-wet peptide/protein array in which peptides could be entrapped within the gel.

The basis of the method relies on the stronger emission of an environmentally sensitive fluorescent probe mixed with the hydrogel, compared with that of the probe in aqueous solution. In this study, the hydrophobic probes ANS (1-anilinonaphthalene-8-sulphonic acid) and DANSen (5-dimethylaminonaphthalene-1-(*N*-2-aminoethyl)sulphonamide), a remarkable blue shift in the emission maximum was observed relative to that in aqueous solution, along with the intensified emission. In contrast, a relatively hydrophilic probe EDANS (5-((2-aminoethyl)amino)naphthalene-1-sulphonic acid) did not show such a strong emission.[75] The results suggest that a hydrophobic domain is created in the present hydrogel, which can be stained with fluorescent probes. Confocal laser scanning microscopy image of the hydrogel containing ANS clearly showed the entangled microfibrils with a diameter of several micrometres in the wet gel state. A very important fact is that these fibrils emit a blue fluorescence due to ANS, indicating that each fibre contains the well-developed hydrophobic core that is able to bind hydrophobic molecules. In a proof-of-principle experiment, lysyl-endopeptidase (LEP) was used. A substrate for LEP was designed that is a hydrophilic pentapeptide (S-S-S-S-K-DanSen) bearing lysine (Lys) and DANSen at the *C*-terminal end. When LEP cleaves a peptide bond between Lys and DANSen, the resultant DANSen shifts from the aqueous cavity of the hydrogel to the hydrophobic domain because of DANSen's strong hydrophobicity, causing the fluorescence increase in the environmentally sensitive dye (**figure 1.11**). Before hydrolysis, the emission maximum is at 540 nm, which is almost comparable to that in aqueous solution (545 nm). After the LEP-catalysed hydrolysis, the emission maximum was shifted to 508 nm, and the emission intensity increased twofold. This colour change can be visually monitored as shown in **figure 1.11**, that is, a pinkish-yellow hydrogel (**figure 1.11 (B)**) turns into a light green (**figure 1.11 (C)**) by LEP addition. In the case of LEP, the emission increase was completed within two hours. The addition of chymotrypsin instead of LEP did not induce such a fluorescence change because chymotrypsin requires an aromatic or hydrophobic side chain on the amino acid (phenylalanine (Phe), tyrosine (Tyr) or tryptophan (Trp)) of the scissile bond. In contrast, a negligible fluorescence change was observed for the reaction in a homogeneous

aqueous solution (that is, in the absence of hydrogel). These results clearly imply that the hydrogel provides a unique reaction medium for sensing enzymatic activity.

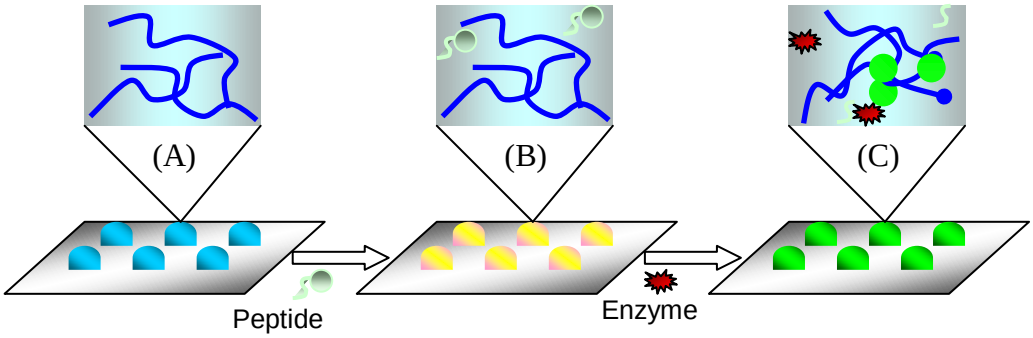


Figure 1.11: Semi-wet peptide/protein chip using a supramolecular hydrogel: (A) hydrogel fibre; (B) Before hydrolysis (the emission maximum is at 540 nm); (C) After the LEP-catalysed hydrolysis (the emission maximum was shifted to 508 nm, and the emission intensity increased by twofold).[75]

1.4. Pro's and Con's of high-throughput protease analysis methods

A large variety of assays are available for the screening for potential protease substrates. The success of an assay critically depends on the design for adaptation and integration of chemistry, screening and structure determination to form a complete process. Thus, synthesis of a diverse set of molecules must be facilitated without compromising the throughput of the screen and analysis. Furthermore, the amount of biological material needed for assaying should be minimal. This point is crucial as most biologically and pharmacologically interesting proteases are available, at least initially, in minute quantities.

For protease profiling, the array techniques with arrays of distinct substrates on glass slides or membranes seem to be a promising technique. On the other hand, in recent years solid-phase synthesis on resin beads has proven a very versatile synthetic technique and many high-yielding organic reactions have been performed on a variety of resins. In biological and biochemical assays on surfaces and on solid supports, the elimination of non-specific binding is crucial and in this regard the polar PEG-based resins such as PEGA have proven particularly useful. For solution-phase assays in titter plates, on surface, and membrane-based assays, non-specific

binding can be minimised by covalent attachment of a coating layer of e.g., PEG or polysaccharide. Glass slides usually present a polar surface and may allow direct attachment of ligands or proteins to the surface.

The screening of discrete fluorogenic peptides in well-based assays requires a considerable amount of compound handling and individual application of conditions to the assay solutions. The mapping of protease activity can very conveniently be carried out in solution using soluble substrate mixtures e.g., a positional scanning format where information of specificity is extracted from one position at a time by either fixing that position in sub-libraries for each position of the peptide, or by omission of one amino acid at a time in each position. The use of all 20 proteinogenic amino acids is not possible due to the sulphur-sulphur bonding that occurs with cysteine when preparing such libraries. The non use of cysteine leads to loss of information regarding other more efficient substrates. Nevertheless, once a superior substrate has been identified this can be synthesised as a fluorescence-quenched version containing a fluorophore and a quencher.

In an even less specific approach, completely randomised dodeca-peptide libraries, which were either N-terminally acetylated or C-terminally biotinylated were subjected to proteolysis.[52-54] N-terminal sequencing of the first library provided the preferred C-terminal to the cleavage site and after removal of any uncleaved and biotinylated material from the second library by streptavidin binding the N-terminal substrate preference could be obtained. Although these techniques are powerful they all inherently assume that there is no additive effect arising from the combination of different sets of amino acid residues within the binding pocket. This may apply in some cases, but is far from general and both additive effects and multiple binding modes have been observed in the interaction between proteolytic enzymes and their substrates.

Array techniques are exquisite tools in the analysis of substrate specificity even for the identification of a particular proteolytic activity in biological sample. In the elegant methods developed by Ellman and co-workers[51,69], small coumarin-containing tetra-peptide constructs were immobilised by arraying the substrates on aldehyde-substituted glass slides using oxime ligation. When treated with a protease each slide showed a different pattern of fluorescent spots depending on the protease

specificity. The method was used to profile the N-terminal specificity of thrombin and results correlated well with solution assays.

Using the same approach, small coumarin substrates were printed with no covalent binding onto polylysine-coated glass slides. Administering the enzyme solution to the slide as a spray of micro-droplets elegantly eliminated the problem of substrate dissolution and mixing on the slide, even if a 100% homogeneous distribution of the micro-droplets is not commensurable. Nevertheless the need for incubation leads to a certain randomness of results, different substrates are cleaved with different ratios and a presupposed loss of 20% signal response on all substrate signals during incubation.[72] The substrates were rather short tetra-peptides that yielded the preferences in sub-sites neighbouring the cleavage site.

The application of solid-phase assays is very attractive to facilitate the above, particularly when screening for unknown enzyme activity or for enzyme inhibitors as this technique provides a complete and statistical picture of the sub-site preferences of the enzyme. In assays of enzymes with delicate function in their chemical transformations, the choice of resin polymer material is crucial for the success of the assay. However, rigorous design of the resin material to provide compatibility with the diffusion and performance of the enzyme reaction is required. The porosity of the resin should match the size of the enzyme investigated. PEG-based resins, such as PEGA 1900 and PEGA 4000, may be used for enzymes with a molecular weight of up to 60,000 and 120,000, respectively. Large protein complexes of more than 150,000 D are best analysed in solution assays e.g., in a well format. Alternatively this type of assay may also be performed on arrays, or on planar surfaces such as membranes or glass slides.

However, a need exists to create localized reaction volumes in an array-based format so as to create a method of rapidly delivering small volumes of fluid to each reaction. Additionally, evaporation effects typically prevent the extreme scale-down of well-plate reactions to nanolitre volumes. An approach to deliver these aims will be discussed in this chapter.

1.5. Dual colour FRET peptide libraries: a novel approach

As seen in the previous section, high throughput enzymatic assays have been widely applied for the generation of data in the area of proteases. Different methodologies are constantly being explored and new assays are constantly being developed to provide improvements in data collection from smaller amounts of compounds and enzymes in a more rapid manner. Different methods using FRET peptide libraries would provide an alternative route in the analysis of proteases. Fluorescence resonance energy transfer (FRET) is a distance-dependent interaction between two dye molecules in which excitation is transferred from a donor molecule to an acceptor molecule *without emission of a photon*. The efficiency of FRET is dependent on the inverse sixth power of the intermolecular separation.[76] Thus, FRET is an important technique for investigating a variety of biological phenomena that produce changes in molecular proximity[77,78]. This approach has been widely used to assay bond-cleavage reactions, particularly in the case of proteolysis of peptides. However, such libraries typically lack internal controls making it difficult to accurately access the relative cleavage of the library members. Fluorescence detection sensitivity is severely compromised by background signals, which may originate from endogenous sample constituents (referred to as autofluorescence) or from unbound or nonspecifically bound probes (referred to as reagent background). Detection of autofluorescence can be minimized either by selecting filters that reduce the transmission or by selecting probes that absorb and emit at longer wavelengths. This constitutes a major barrier in developing fast, low cost and easy to reproduce experimental protocols since it leads to an increase of cost in equipment and/or in fluorescent dyes. For these reasons an easy to prepare dual colour FRET peptide library strategy was designed to reduce the error associated with the usual parallel FRET experiments. The strategy proposed consists of synthesizing a different peptide construct for arrays of peptides using low cost FRET dyes. The construct consist of different peptide sequences, prepared by conventional peptide synthesis with each member labelled with a quencher (i.e. methyl red). Splitting each member of the library into two allows one half to be attached to a quenchable FRET dye (i.e. 5-(6)-carboxyfluoresceine - FAM), and the other half to a non-quenchable FRET dye

(i.e. rhodamine). Once prepared, the two halves of the same peptide sequence would be mixed in equal molar amounts, allowing for each of the two "different" half's of the construct to be printed or immobilised in equal amounts onto a glass slide. This allows an assay to be carried out with an internal control. Analysis at two different emission frequencies gives an initial instant intensity value (t = 0 - **figure 1.12(A)**), where the intensity of the quenched dye is the starting value and the second non-quenched dye gives us what should be a fixed value, which remains constant throughout the assay. Once an enzyme is added, scanning is initiated. The intensity of fluorescence of FAM signal starts increasing as soon as the non-primed peptide containing the quencher is cleaved off, while the non-quenched dye emission intensity signal should remain unchanged (**figure 1.12(B)**), thus allowing kinetics of the protease substrates to be analyzed (**figure 1.12(C)**). The major point in the assay is the control given by rhodamine. Any change in rhodamine intensity provides a clear indication that the considered spot is a false positive and the data should not be considered. By limiting readout parameters to two different wavelength emissions assays, screening of dual colour FRET libraries can be carried out by high-throughput screening, enabling reliable yes-or-no assessments.

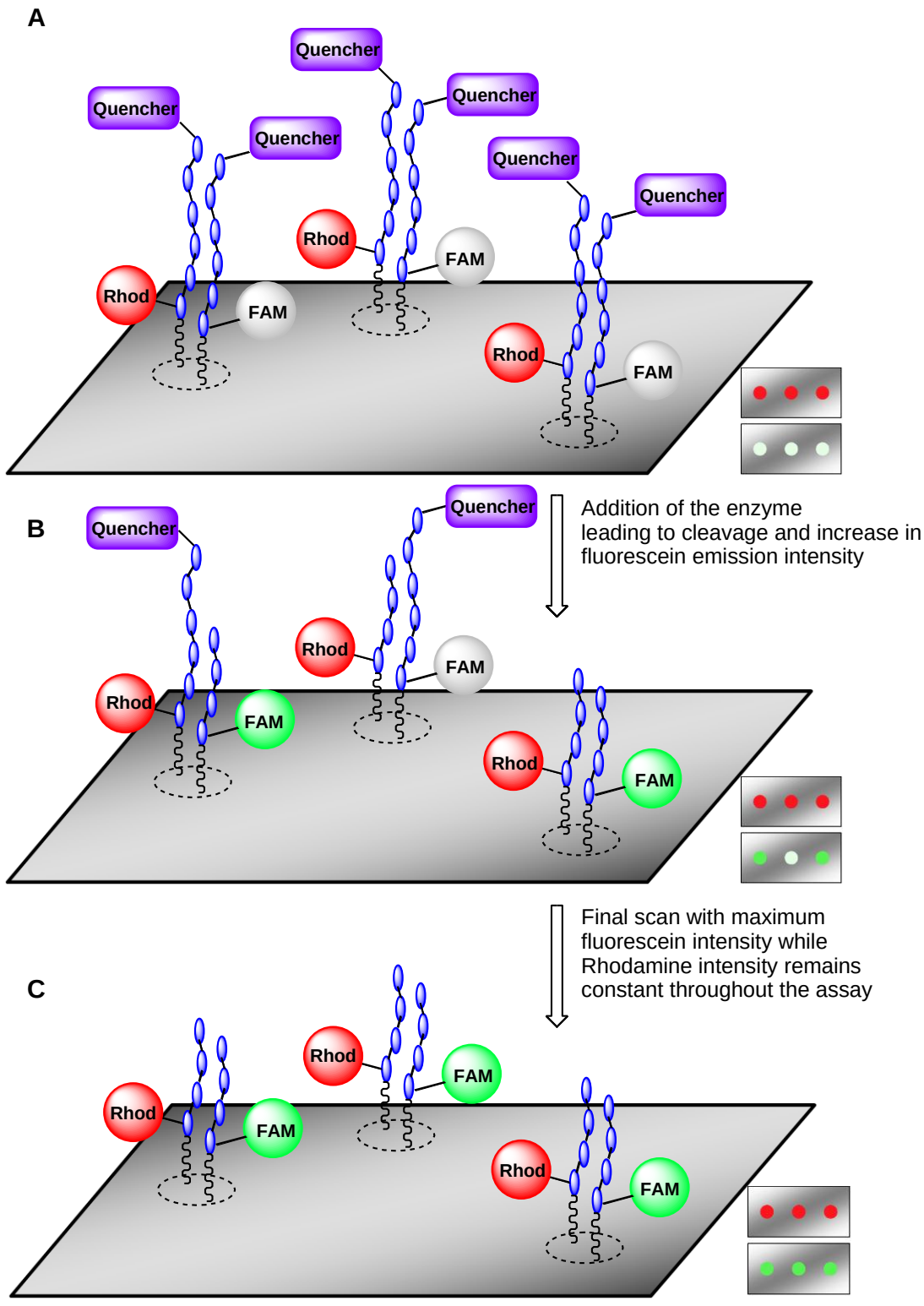


Figure 1.12: Dual color FRET screening assay (A) t = 0, Rhodamine fluorescence detectable, no signal from FAM (quenched); (B) after the addition of enzyme, the most easily cleavable substrates start to show FAM emission, (Rhodamine constant); (C) in the final instance all cleavable substrates have fluorescence emissions for both dyes.

It is possible to dispense the peptides to glass slides in three distinct forms: (i) covalently attaching them by dispensing nanolitre volume spots of the mixed peptide templates onto a glass slide (the composition of each spot would be known without the need for labelling, as long as the printing template is known); (ii) dispensing the peptides in glycerol, were the peptides are suspended in a viscous matrix; (iii) or inside microwells, the peptides being inkjet printed in nanolitre volumes into microwells. Inkjet systems are advanced printing systems, and rely on technology commonly used in desktop printers, where extremely small volumes of solution are deposited without contact between the dispensing tip and substrate. The main advantages of this system are gentle deposition, enabling printing on fragile substrates, and greater control of the quantities of liquid delivered, allowing a wider range of spot sizes to be printed. In the last case a classic well assay would be carried out but on much smaller scale then conventionally possible, while a real time scan would allow collection of kinetic data. In the next chapter these approaches are developed.

1.6. References

(1) Panicker, R.; Chattopadhaya, S.; Yao, S. *Analytica Chimica Acta* **2006**, *556*, 69-79.

(2) Lander, E.; Linton, L.; Birren, B.; Nusbaum, C.; Zody, M.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J.; Mungall, A.; Plumb, R.; Ross, M.; Shownkeen, R.; Sims, S.; Waterston, R.; Wilson, R.; Hillier, L.; McPherson, J.; Marra, M.; Mardis, E.; Fulton, L.; Chinwalla, A.; Pepin, K.; Gish, W.; Chissoe, S.; Wendl, M.; Delehaunty, K.; Miner, T.; Delehaunty, A.; Kramer, J.; Cook, L.; Fulton, R.; Johnson, D.; Minx, P.; Clifton, S.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M. *Nature* **2001**, *409*, 860-921.

(3) Venter, J.; Adams, M.; Myers, E.; Li, P.; Mural, R.; Sutton, G.; Smith, H.; Yandell, M.; Evans, C.; Holt, R.; Gocayne, J.; Amanatides, P.; Ballew, R.; Huson, D.; Wortman, J.; Zhang, Q.; Kodira, C.; Zheng, X.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P.; Zhang, J.; Miklos, G.; Nelson, C.; Broder, S.; Clark, A.; Nadeau, C.; McKusick, V.; Zinder, N.; Levine, A.; Roberts, R.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di Francesco, V.; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T.; Higgins, M.; Ji, R.; Ke, Z.; Ketchum, K.; Lai, Z.; Lei, Y.; Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu,

F.; Merkulov, G.; Milshina, N.; Moore, H.; Naik, A.; Narayan, V.; Neelam, B.; Nusskern, D.; Rusch, D.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C. *Science* **2001**, *291*, 1304-1351.

(4) Kraulis, P. J. *Journal of Applied Crystallography* **1991**, *24*, 946-950.

(5) Merritt, E. A.; Bacon, D. J. In *Macromolecular Crystallography, Pt B* 1997; Vol. 277, p 505-524.

(6) Turk, B. *Nature Reviews Drug Discovery* **2006**, *5*, 785-799.

(7) Zerner, B.; Bender, M. *Journal of the American Chemical Society* **1964**, *86*, 3669-3674.

(8) Rawlings, N.; Morton, F.; Kok, C.; Kong, J.; Barrett, A. *Nucleic Acids Research* **2008**, *36*, D320-D325.

(9) Davies, D. R. *Annual Review of Biophysics and Biophysical Chemistry* **1990**, *19*, 189-215.

(10) Northrop, J. *Journal of General Physiology* **1930**, *13*, 739-766.

(11) Darwin, C. *Insectivorous Plants*; John Murry, London, 1875.

(12) Sorensen, S. *Biochemische Zeitschrift* **1909**, *21*, 201-304.

(13) Newmark, A.; Knowles, J. *Journal of the American Chemical Society* **1975**, *97*, 3557-3559.

(14) Piana, S.; Carloni, P. *Proteins-Structure Function and Genetics* **2000**, *39*, 26-36.

(15) Meek, T. *Catalytic mechanisms of the aspartice proteinases. In Comprehensive Biological Catalysis: A mechanistic Reference*; Academic Press: San Diego, 1998.

(16) Hunkapiller , M. W.; Richards, J. H. *Biochemistry* **1972**, *11*, 2829-2839.

(17) Quinn, D. M.; Sutton, L. D. In *Enzyme Mechanism from Isotope Effects*; Cook, P. F., Ed.; CRC Press: Boca Raton, Fl, 1991, p 73-126.

(18) Cleland, W. W. *Archives of Biochemistry and Biophysics* **2000**, *382*, 1-5.

(19) Hyland, L. J.; Tomaszek, T. A.; Meek, T. D. *Biochemistry* **1991**, *30*, 8454-8463.

(20) Blow, D. *Trends in Biochemical Sciences* **1997**, *22*, 405-408.

(21) Dodson, G.; Wlodawer, A. *Trends in Biochemical Sciences* **1998**, *23*, 347-352.

(22) Rawlings, N. D.; Barrett, A. J. In *Proteolytic Enzymes: Serine and Cysteine Peptidases* 1994; Vol. 244, p 19-61.

(23) Tong, L. *Chemical Reviews* **2002**, *102*, 4609-4626.

(24) Kraut, J. *Annual Review of Biochemistry* **1977**, *46*, 331-358.

(25) Hess, G. P.; McConn, J.; Ku, E.; McConkey, G. *Philosophical Transactions of the Royal Society of London Series B - Biological Sciences* **1970**, *257*, 89-104.

(26) Steitz, T. A.; Shulman, R. G. *Annual Review of Biophysics and Bioengineering* **1982**, *11*, 419-444.

(27) Bender, M. L.; Philipp, M. *Journal of the American Chemical Society* **1973**, *95*, 1665-1666.

(28) Barrett, A.; Salvesen, G. *Proteinase Inhibitors*; Elsevier, Amsterdam, 1986.

(29) Sharma, A.; Padwaldesai, S. R.; Ninjoor, V. *Biochemical and Biophysical Research Communications* **1989**, *159*, 464-471.

(30) Mizuno, K.; Nakamura, T.; Takada, K.; Sakakibara, S.; Matsuo, H. *Biochemical and Biophysical Research Communications* **1987**, *144*, 807-814.

(31) Enenkel, C.; Wolf, D. H. *Journal of Biological Chemistry* **1993**, *268*, 7036-7043.

(32) Menard, R.; Plouffe, C.; Carmona, E.; Storer, A. C.; Krantz, A.; Smith, R. A. *Journal of Cellular Biochemistry* **1994**, 137-137.

(33) Polgar, L. *European Journal of Biochemistry* **1973**, *33*, 104-109.

(34) Keillor, J. W.; Brown, R. S. *Journal of the American Chemical Society* **1991**, *113*, 5114-5116.

(35) Keillor, J. W.; Brown, R. S. *Journal of the American Chemical Society* **1992**, *114*, 7983-7989.

(36) Polgar, L. *European Journal of Biochemistry* **1979**, *98*, 369-374.

(37) Lewis, S. D.; Johnson, F. A.; Shafer, J. A. *Biochemistry* **1976**, *15*, 5009-5017.

(38) Brocklehurst, K. *International Journal of Biochemistry* **1979**, *10*, 259-274.

(39) Foje, K. L.; Hanzlik, R. P. *Biochimica et Biophysica Acta - General Subjects* **1994**, *1201*, 447-453.

(40) Malthouse, J. P. G.; Gamcsik, M. P.; Boyd, A. S. F.; MacKenzie, N. E.; Scott, A. I. *Journal of the American Chemical Society* **1982**, *104*, 6811-6813.

(41) Otto, H.; Schirmeister, T. *Chemical Reviews* **1997**, *97*, 133-171.

(42) Tate, S.; Ohno, A.; Seeram, S.; Hiraga, K.; Oda, K.; Kainosho, M. *Journal of Molecular Biology* **1998**, *282*, 435-446.

(43) Matthews, B. W. *Accounts of Chemical Research* **1988**, *21*, 333-340.

(44) Hangauer, D. G.; Monzingo, A. F.; Matthews, B. W. *Biochemistry* **1984**, *23*, 5730-5741.

(45) Lowe, J.; Stock, D.; Jap, R.; Zwickl, P.; Baumeister, W.; Huber, R. *Science* **1995**, *268*, 533-539.

(46) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chemical Reviews* **2002**, *102*, 4639-4750.

(47) Fujinaga, M.; Cherney, M.; Oyama, H.; Oda, K.; James, M. *Proccedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 3364-3369.

(48) Pinilla, C.; Appel, J. R.; Blanc, P.; Houghten, R. A. *Biotechniques* **1992**, *13*, 901-905.

(49) Rano, T. A.; Timkey, T.; Peterson, E. P.; Rotonda, J.; Nicholson, D. W.; Becker, J. W.; Chapman, K. T.; Thornberry, N. A. *Chemistry & Biology* **1997**, *4*, 149-155.

(50) Thornberry, N. A.; Ranon, T. A.; Pieterson, E. P.; Rasper, D. M.; Timkey, T.; GarciaCalvo, M.; Houtzager, V. M.; Nordstrom, P. A.; Roy, S.; Vaillancourt, J. P.; Chapman, K. T.; Nicholson, D. W. *Journal of Biological Chemistry* **1997**, *272*, 17907-17911.

(51) Backes, B.; Harris, J.; Leonetti, F.; Craik, C.; Ellman, J. *Nature Biotechnology* **2000**, *18*, 187-193.

(52) Turk, B.; Huang, L.; Piro, E.; Cantley, L. *Nature Biotechnology* **2001**, *19*, 661-667.

(53) Turk, B.; Cantley, L. *Methods* **2004**, *32*, 398-405.

(54) Cuerrier, D.; Moldoveanu, T.; Davies, P. *Journal of Biological Chemistry* **2005**, *280*, 40632-40641.

(55) Edman, P. *Nature* **1956**, *177*, 667-668.

(56) Tian, Y.; Sohar, I.; Taylor, J. W.; Lobel, P. *Journal of Biological Chemistry* **2006**, *281*, 6559-6572.

(57) Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F. I. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 3314-3318.

(58) DelNery, E.; Juliano, M. A.; Meldal, M.; Svendsen, I.; Scharfstein, J.; Walmsley, A.; Juliano, L. *Biochemical Journal* **1997**, *323*, 427-433.

(59) Atherton, E.; Brown, E.; Sheppard, R. C.; Rosevear, A. *Journal of the Chemical Society, Chemical Communications* **1981**, 1151-1152.

(60) Rademann, J.; Grotli, M.; Meldal, M.; Bock, K. *Journal of the American Chemical Society* **1999**, *121*, 5459-5466.

(61) Meldal, M.; Svendsen, I.; Juliano, L.; Juliano, M. A.; Del Nery, E.; Scharfstein, J. *Journal of Peptide Science* **1998**, *4*, 83-91.

(62) Rosse, G.; Kueng, E.; Page, M. G. P.; Schauer-Vukasinovic, V.; Giller, T.; Lahm, H. W.; Hunziker, P.; Schlatter, D. *Journal of Combinatorial Chemistry* **2000**, *2*, 461-466.

(63) Meldal, M. *QSAR & Combinatorial Science* **2005**, *24*, 1141-1148.

(64) Diaz-Mochon, J.; Bialy, L.; Bradley, M. *Chemical Communications* **2006**, 3984-3986.

(65) Harris, J.; Mason, D. E.; Li, J.; Burdick, K. W.; Backes, B. J.; Chen, T.; Shipway, A.; Van Heeke, G.; Gough, L.; Ghaemmaghami, A.; Shakib, F.; Debaene, F.; Winssinger, N. *Chemistry & Biology* **2004**, *11*, 1361-1372.

(66) Winssinger, N.; Damoiseaux, R.; Tully, D. C.; Geierstanger, B. H.; Burdick, K.; Harris, J. L. *Chemistry & Biology* **2004**, *11*, 1351-1360.

(67) Winssinger, N.; Ficarro, S.; Schultz, P. G.; Harris, J. L. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11139-11144.

(68) Winssinger, N.; Harris, J. L.; Backes, B. J.; Schultz, P. G. *Angewandte Chemie-International Edition* **2001**, *40*, 3152-3155.

(69) Maly, D. J.; Leonetti, F.; Backes, B. J.; Dauber, D. S.; Harris, J. L.; Craik, C. S.; Ellman, J. A. *Journal of Organic Chemistry* **2002**, *67*, 910-915.

(70) Beythien, J.; Barthelemy, S.; Schneeberger, P.; White, P. D. *Tetrahedron Letters* **2006**, *47*, 3009-3012.

(71) Salisbury, C.; Maly, D.; Ellman, J. *Journal of the American Chemical Society* **2002**, *124*, 14868-14870.

(72) Gosalia, D.; Diamond, S. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 8721-8726.

(73) Gosalia, D.; Salisbury, C.; Ellman, J.; Diamond, S. *Molecular & Cellular Proteomics* **2005**, *4*, 626-636.

(74) Gosalia, D.; Salisbury, C.; Maly, D.; Ellman, J.; Diamond, S. *Proteomics* **2005**, *5*, 1292-1298.

(75) Kiyonaka, S.; Sada, K.; Yoshimura, I.; Shinkai, S.; Kato, N.; Hamachi, I. *Nature Materials* **2004**, *3*, 58-64.

(76) Stryer, L.; Haugland, R. P. *Proceedings of the National Academy of Sciences of the United States of America* **1967**, *58*, 719-726.

(77) Gordon, G. W.; Berry, G.; Liang, X. H.; Levine, B.; Herman, B. *Biophysical Journal* **1998**, *74*, 2702-2713.

(78) Kenworthy, A. K. *Methods* **2001**, *24*, 289-296.

Chapter Two

Novel Dual Colour HT Protease Assays

The purpose of this chapter is to describe the synthesis and application of a dual colour FRET peptide library to analyse protease substrate specificity. HT assays were performed using a number of different platforms.

2. Introduction

The performance of protease assays has been revolutionized by the advent of fluorescence resonance energy transfer (FRET) based assays. The generation of pure, reliable, high-quality substrates is essential for such assays. We need a novel method to generate a small FRET peptide substrate library using microwave assisted SPPS and combinatorial techniques that can provide us with an internal control throughout the course of FRET-based enzymatic assays of this library.

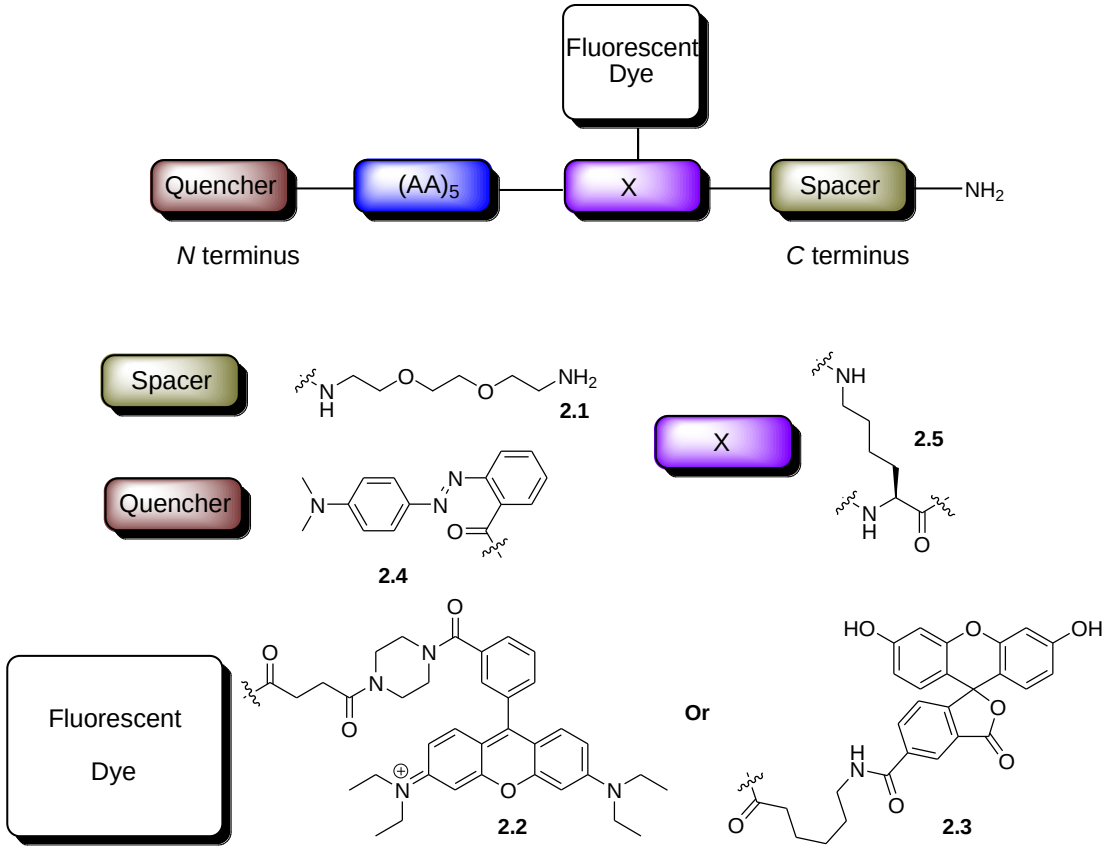
FRET is a phenomenon described by the perturbation in fluorescence emission that occurs with the interaction of two molecules that are in close proximity, with energy transferred from an excited donor molecule (fluorophore) to an acceptor molecule (quencher) being dependent on the Förster radius (R_0), the distance at which 50% efficient energy transfer occurs between these two molecules. Typical Förster radii are in the range of 30–50Å[1]; the FRET efficiency being inversely proportional to the sixth power of the distance between the donor and the acceptor.[2]

Two different approaches have been used to develop FRET protease assays based on minimal effective recognition sequences. Enzymes that have specificity defined by the residues joined to the amino group of the cleavage site of the substrate (P residues) can be assayed using substrates containing a fluorescent leaving group at the carboxy terminus and a quenching group on the other end of the short peptide.

Enzymes with critical carboxy terminus residues at the cleavage site (P' residues) require a second approach. In this case substrates are generated with a fluorophore attached to the amino terminus and a quencher attached to the carboxy terminus (the minimal peptide substrate typically is six residues: three on either side of the cleavage site). A number of fluorophores and quencher pairs have been described for these two approaches.[3,4] In the approach presented here, the peptide had some more modifications when compared to the usual approaches. The substrates were generated with a fluorophore attached to the carboxy terminus and a quencher attached to the amino terminus and the six residue peptides were prepared considering different cleavage lengths (other than three on each side). These modifications were made to determine the importance of residues in the primed and non-primed positions, allowing observation of kinetic differences when different lengths were used (**scheme 2.1**).

2.1. The construct: design considerations

It is critical for the protease assay that the enzymes work when in close proximity to the dye and/or quencher. This is very important as bulky groups may cause disturbance to the normal enzyme active site (the *N*-terminus of the peptide, e.g. the quencher bearing segment, is often not critical for enzyme activity). In the case of the fluorescent dyes, proximity to the peptide chain might cause problems which could be alleviated by incorporating a spacer between the dye and the peptide chain (see **scheme 2.1**).



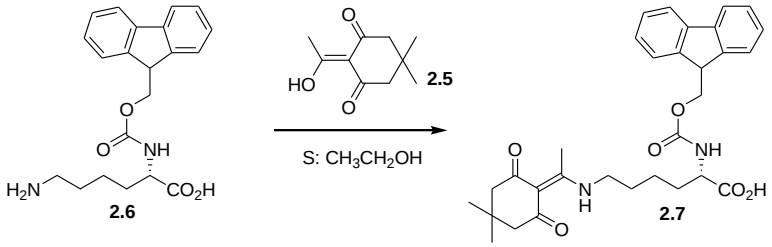
Scheme 2.1: Design of the construct.

2.1.1. Selection of spacer

The design construct required the existence of a reactive group to allow attachment to functionalised glass slides. This led to the use of 2-[2-(2-aminoethoxy)ethoxy]ethanamine (i.e. a spacer containing two polyethyleneglycol units) as a spacer. The primary amine function of this compound provides a perfect reactive group for binding the construct to functionalised glass slides (i.e. epoxy, aldehyde and/or succinic ester functionalised slides). The choice of spacer was also made considering some properties that PEG units could provide to the construct. Linear *C*-chain spacers with more than 10 carbons are often too hydrophobic, while in contrast PEG spacers[5,6] give rise to, less hydrophobic, more buffer/enzyme compatible constructs, are also more flexible, and might be expected to enhance accessibility of proteases to surface bonded constructs.

2.1.2. Selection of first amino acid

In order to prepare our constructs, a starting amino acid residue was used for the attachment of the fluorescent dye and, in a second building phase, the attachment of different amino acids and the quencher was carried out. The first residue had to contain two different protective groups in order to afford a selective deprotection allowing full orthogonality during construct synthesis. The choice relied on the use of an amino acid residue which would embrace the possibility of orthogonally between protective groups. Fmoc-Lys(Dde)-OH[7-9] has been reported to offer orthogonality between the Fmoc/Dde-protective groups and thus, the simple preparation of this doubly protected amino acid, and the simplicity involved in the deprotection required for the Dde group[8,9] appeared to offer a route to the desired material (**scheme 2.2)**.



Scheme 2.2: Preparation of Fmoc-Lys(Dde)-OH.

2.1.3. Choice of fluorophores and quencher

A multicolour labelling experiment requires the deliberate introduction of two or more probes to simultaneously monitor different biochemical functions. Fluorophores with narrow spectral bandwidths are particularly useful in multicolour applications.[10] An ideal combination of dyes for multicolour labelling would exhibit strong absorption at a common excitation wavelength and well-separated emission spectra. Unfortunately, it is not easy to find dyes with the combination of a large extinction coefficient for absorption and a large Stokes shift.

The choice of fluorophores and quencher relied on very simple principles: easy to couple chemistries; low cost starting materials. The basis of the assay: the increase of fluorescence through time for one excitation frequency while scanning at a different frequency remains unchanged. The fluorophores 5(6)-carboxyfluorescein (5(6)-FAM) and rhodamine B, and a quencher methyl red were chosen. Since the main objective is to keep all the assay simple and easy to reproduce.

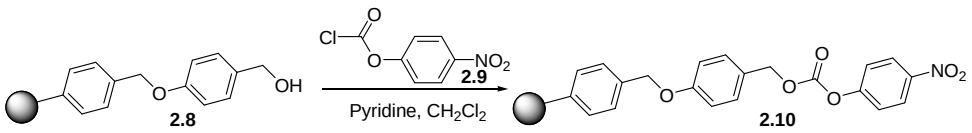
2.2. Synthesis of the library

2.2.1. Preparation of control compounds: selection of solid support.

For successful solid phase peptide synthesis (SPPS) the appropriate choice of solid support is of great importance. Resins used for organic synthesis need to be chemically robust and compatible with a wide range of reaction conditions to allow a diverse range of synthetic steps. To allow easy resin handling and/or automation, the resin also needs to be mechanically robust. In my first approach 2-chloro trityl linker was used in the synthesis of one control compound in order to guarantee the robustness of the resin prior to the library synthesis. Control cleavages were carried out after (1) the attachment of **2.7**, (2) the peptide sequence, the attachment of the quencher, (3) Dde deprotection and after (4) attachment of the two different fluorophores. It was observed that the synthesis proceeded perfectly until the Dde deprotection stage. In this step it was observed that the Dde deprotection was not complete using hydroxylamine•hydrate. Nevertheless, the resin was split and the final coupling of the fluorophores was tried with a very low yield (5(6)-carboxyfluorescein) and no success (rhodamine). Another attempt was carried out by derivatisation of the 5(6)-carboxyfluorescein (**scheme 2.6**) and of rhodamine (**scheme 2.7**) (see **section 2.2.3**).

No success was obtained and a simple empirical study was carried out. An exact amount of 2-chlorotrityl resin and Wang resin with exactly the same reported mesh and cross-linking were weighed, swollen in DMF (the solvent used for all the couplings), and filtered. It was observed that the Wang resin had a volume twice the

one of the 2-chlorotrityl. Considering the possibility that the lack of swellability observed was the cause of the lack of 'accessibility' to the reactive site, it was decided to evaluate the use of the Wang resin. The Wang resin would be reacted with 4-nitrophenylchloroformate (**2.9**) to provide a Wang carbonate resin (**2.9**; **scheme 2.3**) and this could be reacted under the same conditions as previously tried. It was observed that Dde deprotection was complete and the coupling of the flourophores with the spacers was complete (negative results in the ninhydrin tests).



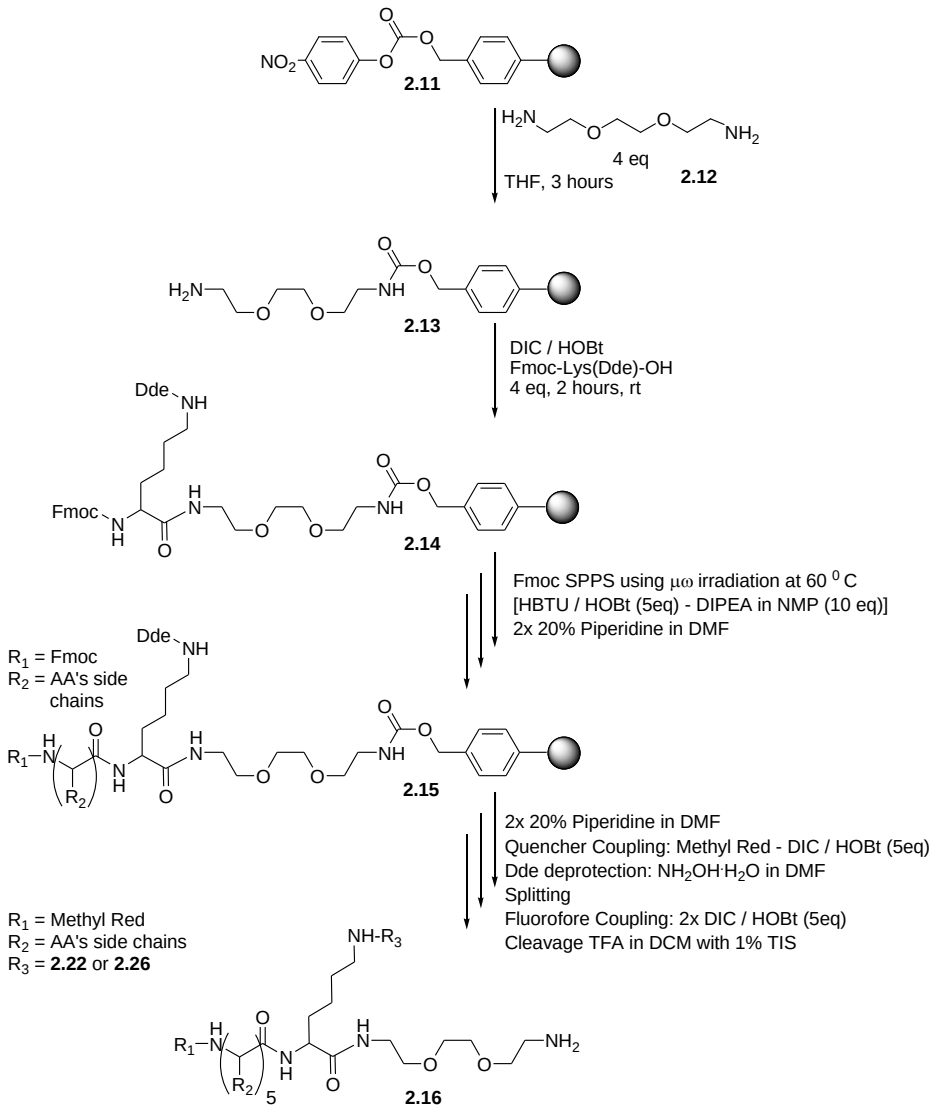
Scheme 2.3: Preparation of Wang carbonate resin.

2.2.2. Automated microwave SPPS

Merrifield first introduced the solid-phase method of peptide synthesis in 1962.[11] Initially, an ester linkage attached the C-terminal amino acid to a polystyrene support. Subsequent amino acids were added by a repetitive protocol of: (1) amino protecting group cleavage by trifluoroacetic acid (TFA); (2) neutralization using triethylamine/dichloromethane; and (3) acylation/amino acid addition, usually using *t*-butoxycarbonyl (Boc) amino acids in the presence of dicyclohexylcarbodiimide and dichloromethane. In 1986, Andrews[12] presented another approach for automated continuous flow peptide synthesis based on the principles laid down by Merrifield, but different from the established technique, paving the way for automated continuous flow peptide synthesis. Firstly, substituting polystyrene with polar polyamide gel supports, enabled polar aprotic solvents such as dimethylformamide to freely permeate the gel media. Second, the use of fluorenylmethoxycarbonyl (Fmoc) amino acid derivatives[13], rather than Boc derivatives, removed the requirement for repeating acid treatments at each deprotection reaction. Third, instead of a simple benzyl ester agent for peptide-resin linkage, the polyamide method uses a range of *t*-Butyl derivatives that can be cleaved with trifluoroacetic acid.[14] Since then several

approaches and developments have occurred but this approach still remains the basis of modern peptide synthesis.

Nowadays the use of microwave energy is widespread for enhancing chemical reactions in organic synthesis, but only recently has it been adopted in SPPS. This has been attributed to concern over well-known side reactions, such as epimerisation and aspartimide formation.[15] Reactions must proceed cleanly and efficiently in order to be successful, and while certain peptide sequences are synthesized relatively easily, some sequences are much more difficult. In some cases, repeated or prolonged reaction times show no improvement in peptide assembly. Optimal reaction conditions require a fully solvated peptide-polymer matrix that allows for efficient reagent penetration. As a peptide is built stepwise on a resin bead, it can form aggregates with itself or with neighboring chains via hydrogen bonding. Microwave energy represents a fast and efficient way to enhance both the deprotection and coupling reactions hindered by aggregation. For this reason and since the purpose behind our approach is to develop a facile methodology from synthesis to analysis, microwave assisted solid-phase peptide synthesis presented itself as the easy and fast tool for the synthesis of the desired peptides. In our case a CEM Liberty peptide synthesizer was used. Protected amino acids with (5 eq. relation to the resin load) where coupled using single couplings of HBTU/HOBt (5 eq.) and DIPEA/NMP (1:4 v/v, corresponding to 10 eq. of DIPEA) under microwave irradiation (40W) for 20 minutes at a temperature of 60^{o}C. The Fmoc protecting group was removed by treatment with piperidine/DMF (1:4 v/v), twice for 8 min. The resin was washed with DMF (3 x 7 mL) and DCM (3 x 7 mL) after each coupling and deprotection step, thus allowing for a simple automated procedure, where no coupling needed to be repeated (**scheme 2.4**). Analysis of the peptides after the sequences were prepared showed clean HPLC traces and the expected masses by MALDI-TOF.

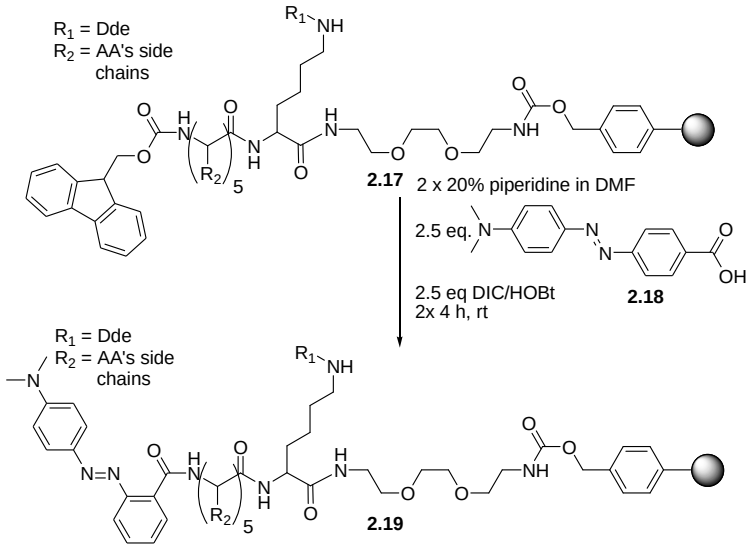


Scheme 2.4: General description of library synthesis.

2.2.3. Attachment of quencher and of FRET dyes

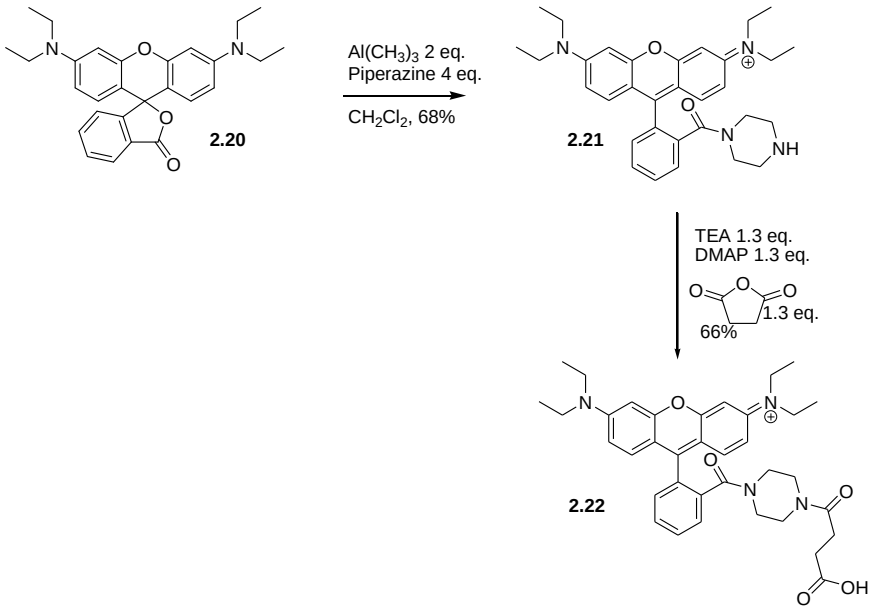
Once the peptides were synthesised the attachment of quencher, split into 2 pools for attachment of the fluorophores was required. As mentioned previously (see **section 2.1.3**) the choice of quencher was methyl red (**2.18**) (**scheme 2.5**). This choice was drawn by its efficiency, in comparison to Dabcyl and its cost. Dabcyl is often used in the common belief that this quencher is very easy to couple. The coupling of the quencher was tested using four different procedures: HATU, HBTU, PyBop and DIC/HOBt. In all cases an increase in the impurities cleaved from the

resin were observed by HPLC, but less in the case of DIC/HOBt which also gave the best conversion overall after one coupling cycle (approx. 85%, with some samples presenting complete conversion). It was verified that the coupling was in some cases not complete after a single coupling, but all peptides were successfully coupled with methyl red after two coupling cycles. Each cycle was carried out for two hours at room temperature using DIC/HOBt as coupling reagents (**Scheme 2.5**).



Scheme 2.5 – Coupling of the quencher.

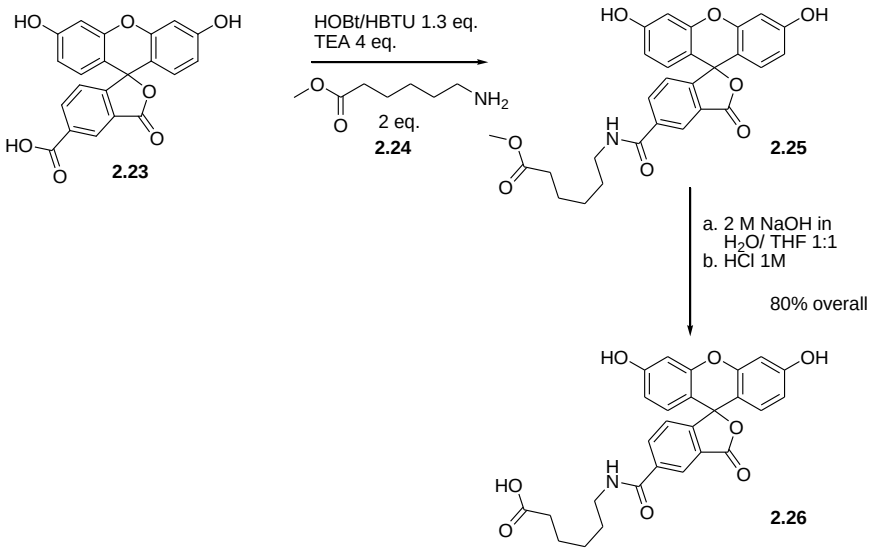
As referred to previously (see **section 2.2.1**), the two selected fluorophores were derivatised in order to attempt to solve the problem of coupling the dyes onto 2-chlorotrityl resin. Since the coupling of the fluorophores had constituted a problem the couplings of these were tested using the Wang resin. Since no problems were found coupling the derivatised fluorophores there was no reason not to use compounds **2.22** (**scheme 2.6**) and **2.26** (**scheme 2.7**) which had been prepared in a reasonable amount previously.



Scheme 2.6 – Synthesis of the rhodamine B derivative used[16].

For the synthesis of the rhodamine B derivative[16] (**2.22**), lactone **2.20** was exposed to 4 equiv of piperazine and 2 equiv of $AlMe_3$ in CH_2Cl_2 at reflux resulting in a clean conversion to tertiary amide **2.21**. The secondary amine of **2.21** was elaborated to acid **2.22** by reaction with succinic anhydride, NEt_3, and DMAP to provide the common carboxylic acid group used in peptide coupling reactions (**scheme 2.6**).

For the synthesis of the FAM derivative (**2.26**), lactone **2.23** was coupled to **2.24** using HOBt/HBTU (1.3 eq.) in the presence of TEA. Once prepared **2.25** was subject of a saponification and **2.26** was extracted without the need for further purification (**scheme 2.7**).



Scheme 2.7 – Synthesis of the FAM derivative used.

In the final step, the fluorophore derivatives were then coupled using DIC/HOBt (2x 5eq) producing **2.16** (**scheme 2.5**).

2.3. Properties Considered for Selection of Proteases

Amino acid sequence data is now available for over 450 peptidases (endopeptidases and exopeptidases) found in over 1400 organisms (bacteria, archaea, archezoa, protozoa, fungi, plants, animals, and viruses), and they have been organized into evolutionary families and clans by Rawlings and Barrett.[17] This effort led to the development of the MEROPS[18] database (http://merops.sanger.ac.uk), which now includes a frequently updated listing of all peptidase sequences.

The choice of proteases was based on ease of handling, the sizes of the enzymes, availability and costs. The size of the enzyme, although not such an important property in solution phase screening, could be crucial when applied to surface screening methodologies were arrays of molecules are highly dense. For this reason, enzymes of a molecular weight, lower than 40 kDalton, were focussed upon. Many proteases are relatively non-specific. Summarised in table 2.1 are some high-lights of those chosen that guided library synthesis (i.e. the selection of peptide sequences – Section 2.4).

Table 2.1 - Selected information about the enzymes that were to be employed when assaying the peptide library.

Protease	MW (kDa)	Catalytic type	Specificity
Subtilisin	27	Serine	Subtilisin presents some specificity in subsites S_1 and S_4. In S_1 non-β-branched hydrophobic side chains are preferred. The S_4 subsite strongly prefers hydrophobic side chains.
Chymopapain	23	Cysteine	Chymopapain is considered to possess a fairly broad specificity but subsite S_2 prefers bulky non-polar side chains and a weak preference for Lys in P_1 has been observed.
Thermolysin	35	Metallo	Thermolysin prefers to cut the scissile bond at the *N*-terminal side of Leu, Phe, Ile, and Val and a hydrophobic residue is preferred at the P_1 position.

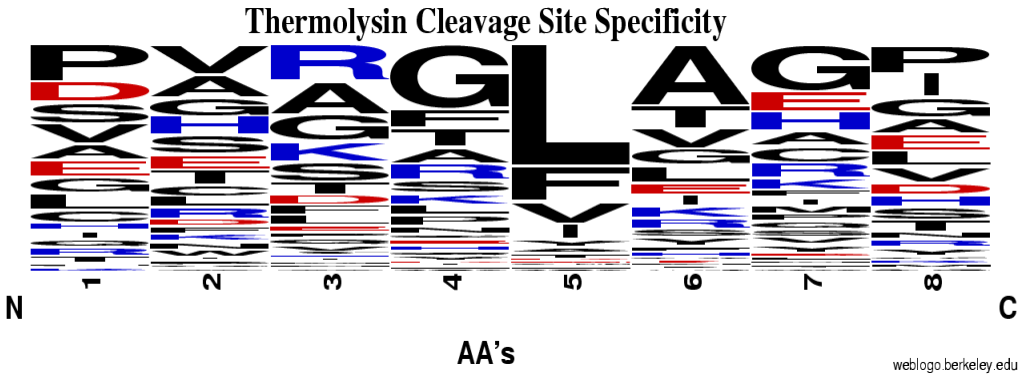
2.4. Selection of Peptide Sequences

Each new update of the MEROPS database adds new members and families, but also specificity matrixes of preferential positions of amino acid residues collected from reported cleaved substrates for each member of each family. The sequences prepared were based on referenced substrates and cleavage site specificity matrixes, more specifically in the cases of thermolysin[19,20] and subtilisin[21-23]. The sequence selection for chymopapain, were based on the reported cleavable substrates determined in Edinburgh.[24] Also, six sequences were designed considering the fact that none of them had been reported as substrates for the enzymes under study, and could provide a set of controls (**table 2.2)**.

Table 2.2 – Selected peptides sequences for each enzyme.

Thermolysin	Chymopapain	Subtilisin	Controls
Y-G-G-F-L-K	S-A-L-F-I-K	V-N-Q-H-L-K	P-I-Q-Q-H-K
G-F-F-Y-Y-K	A-L-F-I-A-K	E-A-L-Y-L-K	H-I-P-P-Y-K
E-A-L-Y-L-K	S-L-F-I-A-K	F-F-Y-T-P-K	P-I-H-Y-Q-K
Y-Q-L-E-N-K	S-A-F-I-A-K	A-L-Y-L-V-K	H-P-I-Q-Q-K
Y-N-Q-F-F-K	S-A-L-I-A-K	Q-H-L-C-G-K	I-H-P-Q-P-K
G-S-H-L-V-K	S-A-L-F-L-K	L-C-G-S-H-K	P-I-P-I-S-K

In the case of thermolysin the selected sequences were based on 84 reported cleavages.[19,20] As observed in **figure 2.1** this enzyme presents a selective preference for substrates containing leucine at the P_1' position, but also some specificity is registered for phenylalanine and valine.



-/-/-/-$\downarrow$$L_{FV}/_{A}$/-/-

Figure 2.1 - Diagrammatic representation of the specificity preference, larger and smaller, upper- or lower-case letters are used to give an intuitive feel for the strength of the specificity preference in each of the subsites P_1 – P_1' for 84 reported thermolysin substrates (P_1 = 4 and P'_1 = 5).

In the case of subtilisin the selected sequences were based on 24 reported cleavages, although no specificity preference could be clearly observed (**figure 2.2**). Nevertheless the substrates selected for the assays should be cleavable, based on reported results.[21-23]

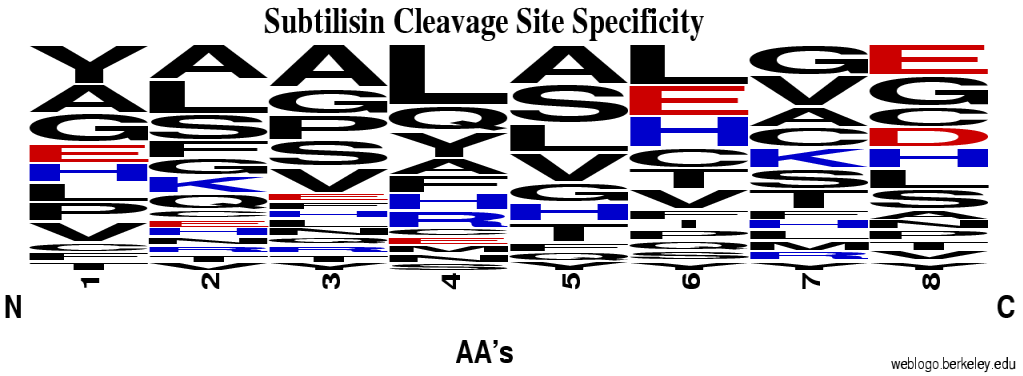
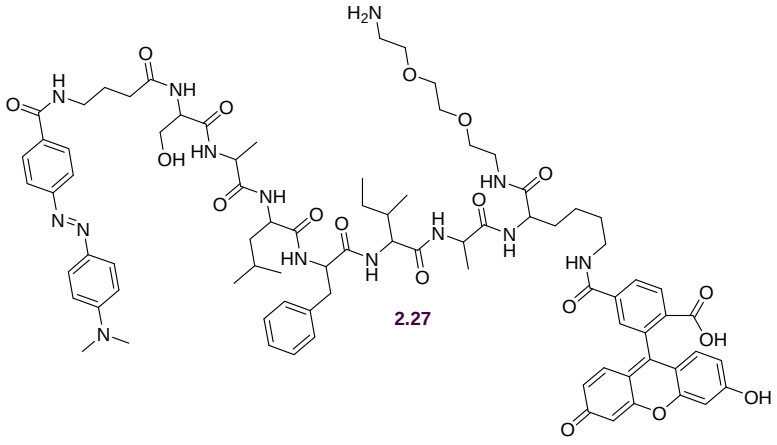


Figure 2.2 - Diagrammatic representation of the specificity preference, larger and smaller, upper- or lower-case letters are used to give an intuitive feel for the strength of the specificity preference in each of the subsites P_1 – P_1' for 24 identified subtilisin substrates (P_1 = 4 and P'_1 = 5).

In the case of chymopapain few reports of substrate specificity have been recorded at the present time, and the choice of substrate relied on an identified substrate recorded in-house.[24] In the reported case the cleavage site had not been determined, and the choice of substrates was made in order to attempt to identify the substrate's P_1 and P_1' positions.



Scheme 2.8 – Identified substrate for Chymopapain (Sequence Dabcyl-S-A-L-F-I-A-K-FAM).[24]

From substrate **2.27** (**scheme 2.8**) all possible P_1 and P_1' combinations were taken into account and considered to identify the substrate's cleavage site. These were

prepared in a manner to obtain some cleavable substrates, due to the similarity to substrate **2.27.**

Table 2.3 – Representation of the possible cleavage positions with respect to the prepared substrates and the specificity of the enzyme for the possible P_1 and P_1' positions.

P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'
			S	A	L	F	I
			S	A	F	I	A
			S	A	L	I	A
			S	A	L	F	L

P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'
		S	A	L	F	I	
			A	L	F	I	A
		S	A	L	I	A	
		S	A	L	F	L	

P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'
	S	A	L	F	I		
		A	L	F	I	A	
		S	L	F	I	A	
	S	A	L	F	L		

P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'
S	A	L	F	I			
	A	L	F	I	A		
	S	L	F	I	A		
	S	A	F	I	A		

P_4	P_3	P_2	P_1	P_1'	P_2'	P_3'	P_4'
S	L	F	I	A			
S	L	F	I	A			
S	A	F	I	A			
S	A	L	I	A			

From the 20 natural occurring amino acids 16 were used to construct the library, excluded were arginine, asparagine, methionine, tryptophan and lysine.

Arginine and lysine were excluded due to the fact that attachment to the supports was going to occur by reaction of primary amine of the construct to amine-reactive functionalised surfaces and the use of these amino acids would cause non desired attachment. Methionine and tryptophan were not used due to: the fact that for the selected enzymes, very few specific substrates have been reported where these amino acids are pivotal either in primed or non-primed positions.[18]

Side chain functional groups of the amino acids used were protected with acid labile protective groups: cysteine, glutamine and histidine with a trityl group; aspartic acid and glutamic acid with a *tert*-butyl ester group; and serine, threonine and tyrosine with a *tert*-butyl ether group. This allowed removal of the protective groups with acid treatment and concomitant release of the construct from the resin, in one final step.

A total of 24 unique sequences were chosen (**Table 2.3**), which after splitting and labelling would provide a small library of 48 hexapeptides.

2.5. HT assays: assessment of different platforms

It is possible to dispense peptides on to glass slides in three distinct forms: (1) covalently attaching them by dispensing nanolitre volume spots of the mixed peptide templates onto a glass slide (the composition of each spot would be known without the need for labelling, as long as the printing "grid reference" is known);(2) dispensing the peptides in glycerol, where the peptides are suspended in a non-volatile matrix; (3) the peptides could be inkjet printed in nanolitre volumes into microwells made from a polystyrene coated glass slide. In this case a classic well assay would be carried out but on a much smaller scale then conventionally done. A real time scan would allow collection of data over time.

In order to evaluate the best technique in which to produce reliable results using the smallest amount of sample several methodologies were investigated. In order to use the minimum amount of sample microarray technologies were used to print the peptide substrates either when attached to the slides or in the "gel based assays" or in microwells. Microarray printing ensured delivery of nano- and picolitre volumes of materials onto a substrate in a defined pattern with speed, reproducibility

and at high density. This capability is aided by the development of robotic instrumentation, and currently there are a huge range of different technologies available. The printing systems will be divided into two main technologies: contact-printing (also known as pin-printing) and ink-jet based deposition (also called reagent jetting) technologies and both were used in this project.

2.5.1. Glycerol droplets & spray assay

In order to identify which substrates would be cleavable an initial approach, based on the spray assays of Gosalia [25-28] were carried out. The library was printed (by contact printing) in different formats. In one format, the constructs were printed in glycerol, a spray of water was used to "solvate" the printed droplets and afterwards the enzyme was sprayed and subsequently incubated and scanned. In the initial phase of the process some response in terms of an increase in fluorescent signals on certain spots was observable but with a problem: only some spots of the 4x4 printed pattern of substrates showed such increase (**figure 2.3**). Also, after a few hours of incubation the spots tended to "slide" around the glass surface which made identification of each substrate difficult (**figure 2.3**). Two main reasons associated with such problems were identifiable: the spraying system was not homogeneous; the printed spots during the incubation periods were kept in a saturated atmosphere and this leads to condensation onto the droplets and movement of the drops over the surface. Attempts were made to make more uniform and reliable assays. Some assays were attempted in a non saturated atmosphere but, as expected, this resulted only in dry droplets with absolutely no signal. Attempts were carried out in order to analyse the influence of the nature of the glass surface without the attachment of the substrates. In this assay, the glass slides were coated with (tridecafluoro-1, 1, 2, 2-tetrafluorooctyl) dimethylchlorosilane in order to observe the influence of a hydrophobic surface. However, here the droplets migrated even more easily than before. This at first seemed to be another discouraging result but, as will be presented on section 2., this observation proved to be crucial.

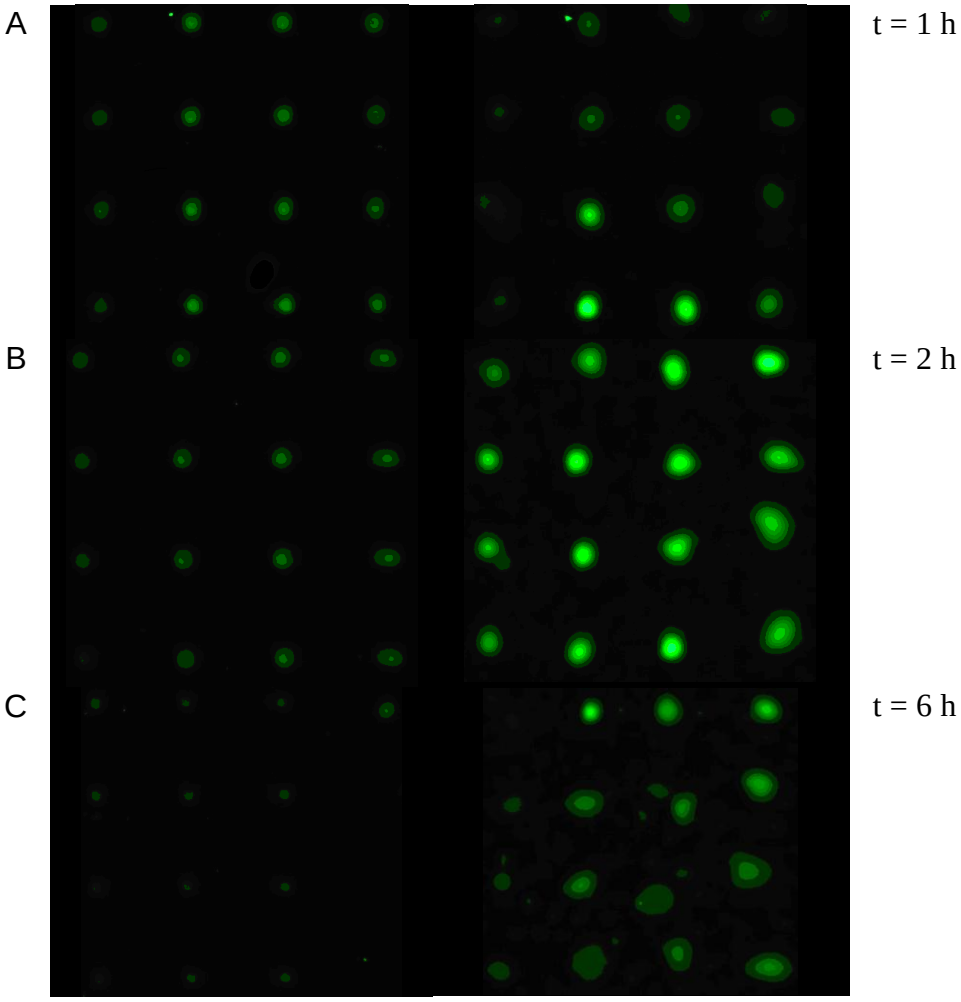
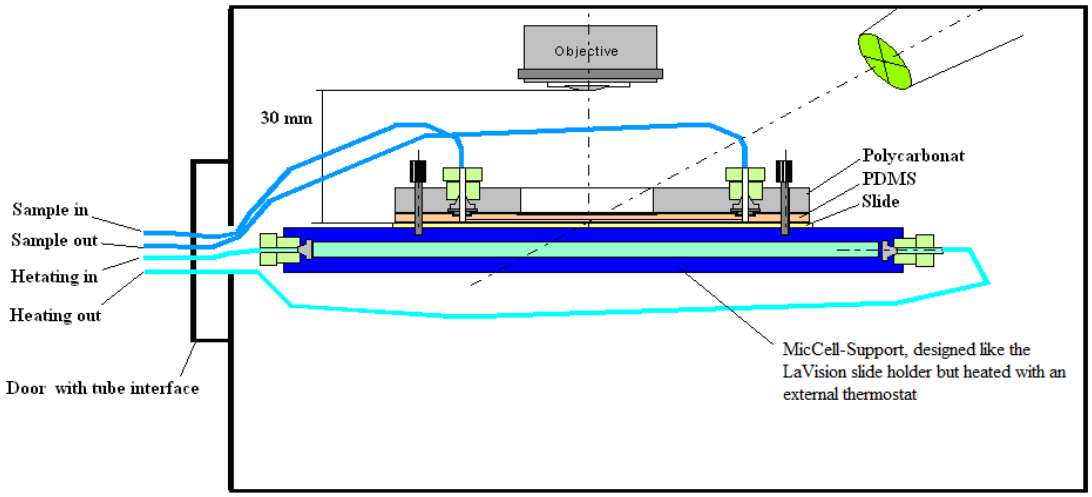


Figure 2.3 – Scanning of the glycerol droplets using thermolysin and the substrate sequence E-A-L-Y-L-K printed in a 4 x 4 pattern. On the left, the rhodamine filter signals as controls showed a decrease in fluorescence with time. On the right, from A to C it is visible the "sliding" of the droplets and loss in fluorescence signal.

2.5.2. Microarray substrate bonded assays

Microarray assays with the constructs bonded to the glass slides gave a solution to the problems identified in the glycerol-based assays. These were carried out using three different types of functionalised glass slides: epoxy, *N*-hydroxy succinimide and aldehyde. Another difference in this procedure was the fact that the active ester functionalized slides were coated with a proprietary 3-D surface chemistry comprised of a long-chain, hydrophilic polymer containing amine-reactive groups. This polymer is covalently cross-linked to the surface of the slide points and

displays the immobilized constructs in a more three-dimensional manner. In both aldehyde and epoxy functionalised slides immobilization gives a more tri-dimensional orientation of the constructs. In order to obtain a real-time scanning without incubation and washing steps, a platform for the enzymatic assays was used (**figure 2.4**). The MicCell-Support consists of an externally heated chamber, where the enzyme solution under study can be introduced and monitored in real time using a LaVision scanner. This allows control of the temperature of the assays in real-time monitoring, as well as a flow control of the enzyme solution.



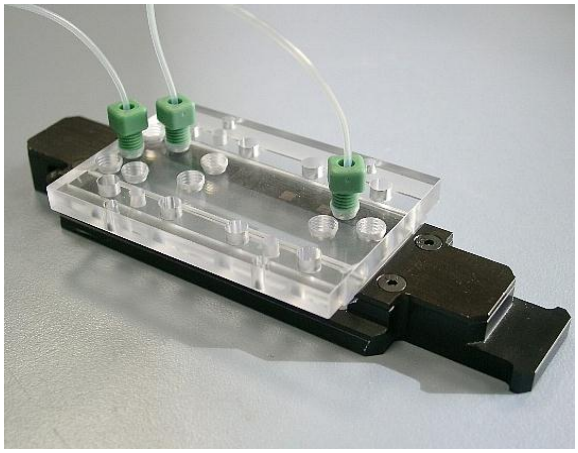
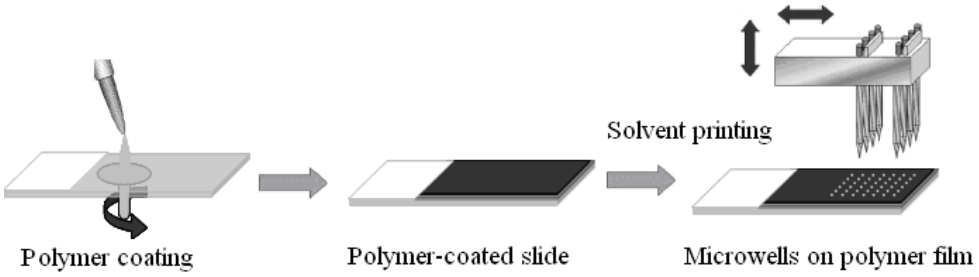


Figure 2.4 - MicCell-Support slide holder used in the construct-bonded assays. On top a schematic representation of the scanning using the support. Below, the slide holder.

Although during the assays carried out using the glycerol spray cleavage had been observed, this was not the case with the attached constructs in the microarray assays. In a first instance the flow of the enzyme solution was altered and assays were run without flow, but the results were the same and no change of fluorescence occurred. The concentration of the enzyme solution and even the use of another enzyme (firstly thermolysin and secondly subtilisin) were evaluated in order to prove that the problem was not with the enzymes but in the format of the assay. The attachment of the constructs clearly limited the mobility of the enzymes and didn't allow cleavage to occur.

2.5.3. Microwells assay

Microwell technology was regarded as a possible way to obtain reliable results.[29] In the applications carried out there were several differences from the microwell applications usually carried out in enzymatic assays. In most reported publications substrates or enzymes are usually attached to the surface of the prepared microwells. This would not be the case regarding our applications, since neither of these would be attached. Here we used the reported fabrication of arrays of 3D microwells by contact printing of a mixture of solvents (acetophenone/ethyl acetate; 2:1 v/v) upon polymer layered coated substrate consisting of polystyrene (PS) (Mn~140,000 and Mw~230 kDa) using a microarray printer.[29]



Scheme 2.9 - General process of polymer microwell fabrication by contact printing[29] (reproduced with permission).

This approach offers access to a highly controllable variety of microwells in an array type format, with high densities of wells, each of which is individually addressable by the printer. The general approach is shown in scheme 2.9 in which a contact-based microarrayer equipped with solid pins was used to deposit organic solvent onto polymer coated glass slides, causing local dissolution and microwell formation. For our application the microwells were generated using a combination of 24 solid pins, and the solvent mixture printed (acetophenone/ethyl acetate; 2:1 v/v) (**scheme 2.9**). Subsequently the peptides were printed into the wells. This allows for the pin that produces the well to print the peptide exactly inside the well. Nevertheless, this leads to some geometry inconsistencies due to the positional accuracy required. When inkjet printing the enzymes it was observed that in some cases the enzyme printing occurred outside of the microwell. In a first instance these geometry inconsistencies could be resolved by using a single solid pin to produce the microwells (**figure 2.5**).

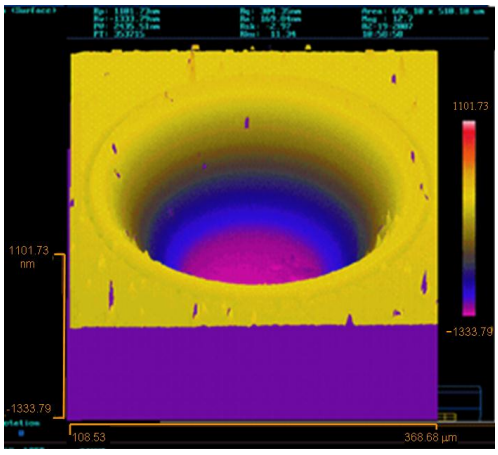


Figure 2.5 - A single microwell fabricated on a PS film (2.4 µm thickness) by stamping acetophenone/ethyl acetate 8 times with a 150 µm diameter solid pin (K2783) (reproduced with permission).[29]

This would lead to a very long printing operation with all the possibilities of change in volume either by evaporation or by swelling. An inaccurate knowledge of the volume due to these phenomena limits the possibility of obtaining quantitative results without largely increasing the errors associated. Another possibility would be the use of the same 24 pins and, by contact printing, print the enzyme into the microwells. In this case the use of the pins that lead to different printing volumes would not permit an accurate volume determination of constructs and enzyme in each of the

microwells. Regardless of these issues one thing was achievable: a qualitative evaluation of the library was possible and confirmation of an increase of fluorescence of some substrates once submitted to printing with the enzymes and good control of the intensity of the control signal were obtained and this confirmed the results of the assays carried out using the glycerol droplets and the reliability of the design of the constructs.

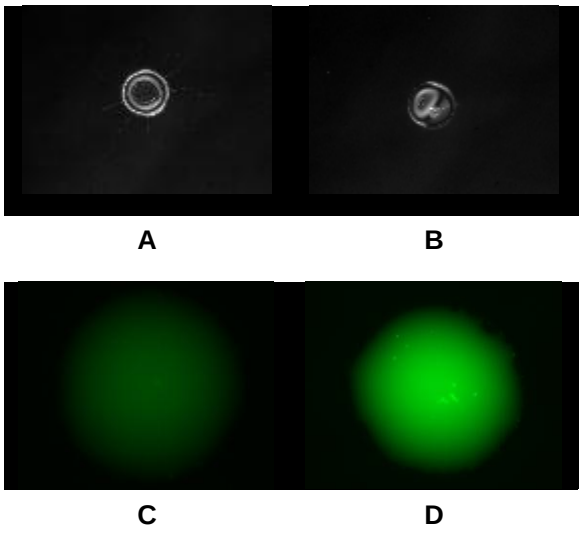


Figure 2.6 – (A) Microwell prior to printing of substrate; (B) Microwell after printing of substrate; (C) Fluorescein signal of microwell just after ink-jet printing of enzyme; (D) Fluorescein signal of microwell after 2 hours of printing. Subtilisin was used and the construct sequence here identified was V-N-Q-H-L-K Images were captured using a Nikon microscope equipped with the Pathfinder software (IMSTAR S.A., Paris, France).

In a different procedure, the microwells were all prepared using a single pin, to ensure consistency and the substrates and enzymes inkjet printed. Once again, there were some difficulties observed, specifically in the inkjet printing of the substrates. This could be related to the different hydrophobic ties of the chains. Once the voltage of the nozzle is set all the peptides are printed with the same voltage, and in some cases, this lead to a printing outside the wells. Another issue with the assays was the increase of background intensity while scanning. The polystyrene coating was responsible for such increase, which was in part expected, and this was resolved using a Nikon microscope equipped with the Pathfinder software (IMSTAR S.A., Paris, France) for data collection. Nevertheless this equipment requires to collect data for each microwell individually, leading to a long time scanning from the first well until the last, thus limiting any possibility for kinetic studies (for a slide containing 72 microwells a scan of each well takes a total time of around 45 minutes). For these

reasons, and since positive results regarding the cleavage had been identified, a move on to a different assay method was carried out and this is described in the following section.

2.5.4. Hydrophobic Masked Slides Assays

In order to eliminate the printing effect that was observed during the previous assays another method was used. This approach consisted of the printing of substrates and enzymes by inkjet printing into an in-house prepared glass slide. This is similar to many of the tested assays reported in the literature except for the fact that the glass slide was firstly subjected to a sucrose printing followed by a coating with an hydrophobic substance (i.e. (tridecafluoro-1, 1, 2, 2-tetrafluorooctyl) dimethylchlorosilane) and a final washing with water (3x) and acetone (3x) to provide a slide with an hydrophobic surface but with defined hydrophilic areas (**figure 2.7**).

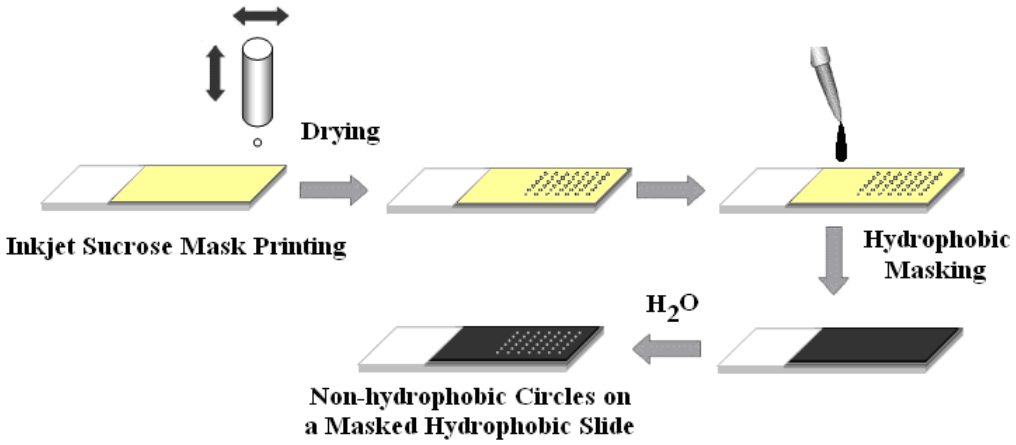


Figure 2.7 – Preparation of the in-house masked slide using (tridecafluoro-1, 1, 2, 2-tetrafluorooctyl) dimethylchlorosilane to create a hydrophobic masked slide with hydrophilic areas.

Once the masked slide was prepared the peptide library was printed out, using a known amount of each peptide dissolved in a mixture of PBS/ glycerol/ DMSO (in 9:9:1 ratio respectively). The use of DMSO was required to guarantee the

solubilisation of the library, while the PBS was used in order to reduce the viscosity. The use of glycerol alone had led to printing problems visible in figure 2.8 (shifting to the right, **figure 2.8**).

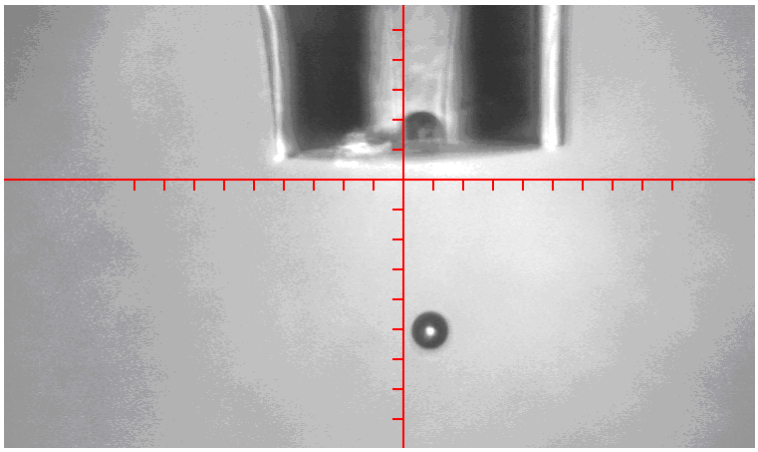
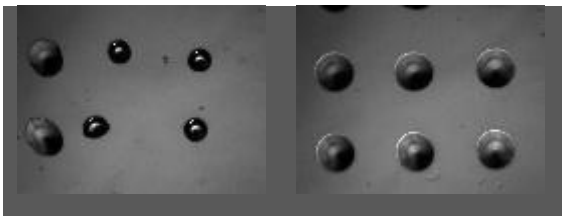


Figure 2.8 – Shifting to the right while inkjet printing glycerol.

Due to the way the voltage of the nozzle and the different peptides interacted, it was required that the masking eliminated any printing incongruence (**figure 2.9**). This was the case and, in figure 2.9 (A), it is observed that the droplets printed on the hydrophobic surface tend to "slide" around (as described in **section 2.5.1**) while in figure 2.9 (B) it is clear that the printed peptides appear in a circular and more spread out manner, thus making the library an easier "target" when inkjet printing the enzymes (**figure 2.9**).



A B

Figure 2.9 – (A) On the right two drops printed inside the hydrophilic area next to four droplets on top of the hydrophobic mask; (B) six samples printed inside the hydrophilic areas.

In order to evaluate the reliability of the dual colour control an assay was first carried out by printing 20 droplets of the peptides and printing in the odd positions 20 droplets of buffer (the specific buffer for the enzyme), and on the even positions

20 droplets of enzyme solution. It can be clearly observed that in a first instance inside the green area on the left the intensity of all the signals remains the same, but after 3 hours, inside the green area on the right the even positions shows an increase in signal intensity, due to the substrate being cleaved. In the red box it is easy to see that no significant change in intensity for the control occurred thus providing a reliable control for these assays (**figure 2.10**).

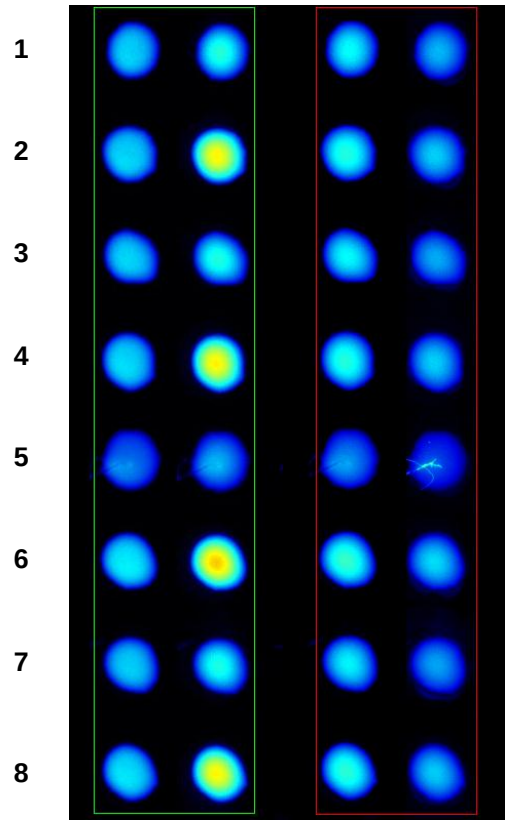


Figure 2.10 – On the left in the green area the starting an final scans for the FAM filter of the substrate V-N-Q-H-L-K using subtilisin, while on the odd positions was printed the same volume of buffer as in the even positions were the enzyme was printed. On the right the same area but using the Cy3 filter, not showing any variation of intensity.

In another assay, data was collected in different time intervals: starting scan; after 15 minutes; after 1 hour; after 2 hours; after 6 hours and after 24 hours (**figure 2.11**). It is clear that an increase in signal intensity for the fluorophore. In the case of the substrate V-N-Q-H-L-K (subtilisin substrate), the intensity of the fluorescence signal doubled from the starting scan until the final scan, as was the case for other substrates for thermolysin and chymopapain (assays being carried out at room temperature; **figure 2.11**).

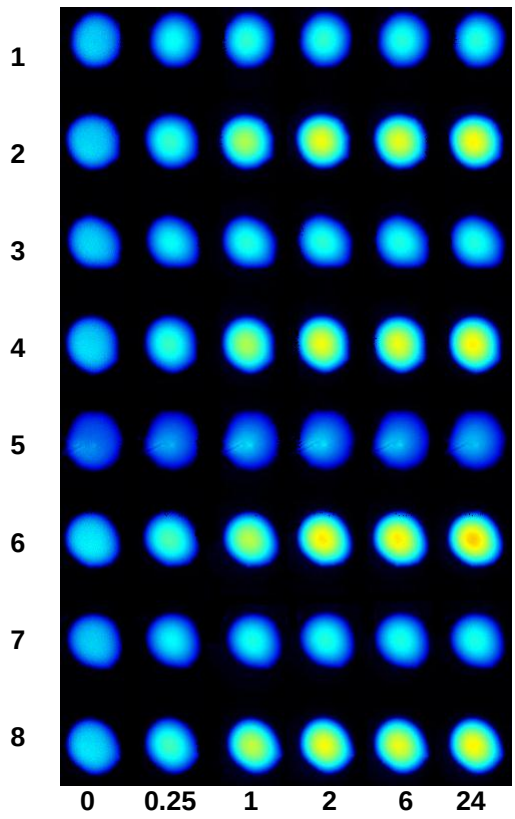


Figure 2.11 – On the odd positions was printed buffer in the same number of drops as in the even positions were the enzyme was printed. From left to right, the increase in fluorescein signal through time. Substrate V-N-Q-H-L-K using subtilisin.

In the case of thermolysin which has an optimal temperature for activity between 60ºC and 70ºC, registering cleavage in a time interval of 6 hours was very promising (**figure 2.12**). One observation that was very elucidative was the fact that the peptide with the sequence E-A-L-Y-L-K, presumed to be substrate for both subtilisin and thermolysin, showed very different results when used with the different enzymes. **Figure 2.12** shows the results for this substrate for both these enzymes: in both the plots series from 1 to 4 represent the intensity in the fluorescein signal vs. time; series 5 to 8 represent the intensity for Cy3 vs. time. On the bottom can be clearly observed that intensity increase in series 1 to 4, while series 5 to 8 show little variation (**figure 2.12**).

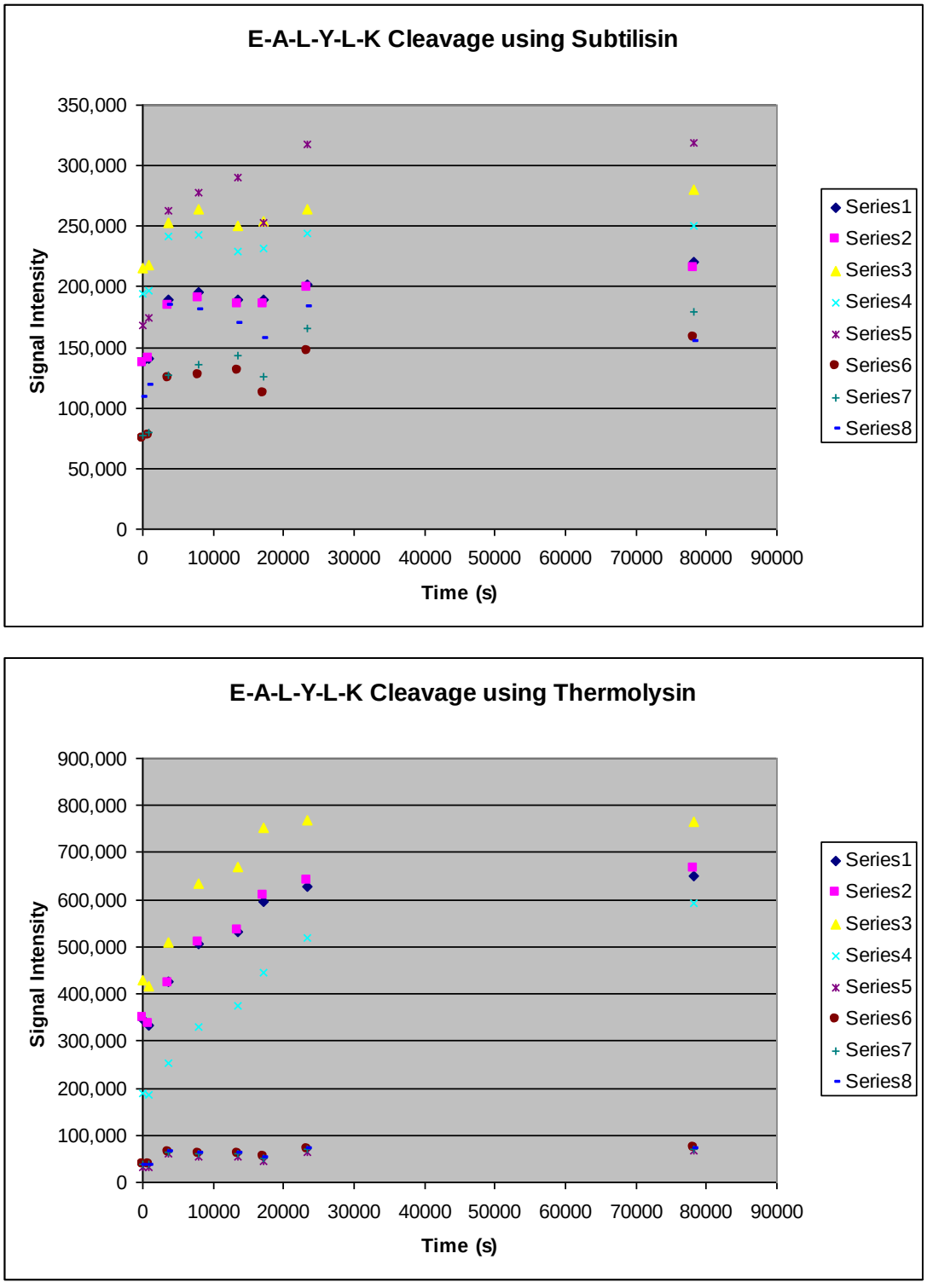


Figure 2.12 – Evaluation of E-A-L-Y-L-K with Subtilisin (top) and Thermolysin (bottom). Series from 1 to 4 represent the intensity in the fluorescein signal vs. time, in four different assays. Series 5 to 8 represent the intensity for Cy3 vs. time, in four different assays.

This is the response that was expected. The assays were optimized to use the smallest amount of enzyme, with a response time around 6 hours representing an appropriate interval of time for data collection. For the different enzymes different amounts were found to be optimal (**table 2.4**).

Table 2.4 – Enzyme units used after optimization.

Protease	Units per μl	Buffer
Subtilisin	5×10^{-2}	0.1 M Tris pH 8.0; 5 mM $CaCl_2$
Chymopapain	1	0.2 M NaCl, 6. mmol Mercaptoethanol; 5 mmol EDTA; 50 mmol *L*-Cysteine hydrochloride hydrate
Thermolysin	1×10^{-3}	0.1 M Tris pH 8.0; 10 mM $CaCl_2$

In the cases of chymopapain and of subtilisin, only one substrate of the six prepared for each of these enzymes respectively showed positive results (**figure 2.13** overleaf). The control sequences presented no cleavage from any of the enzymes.

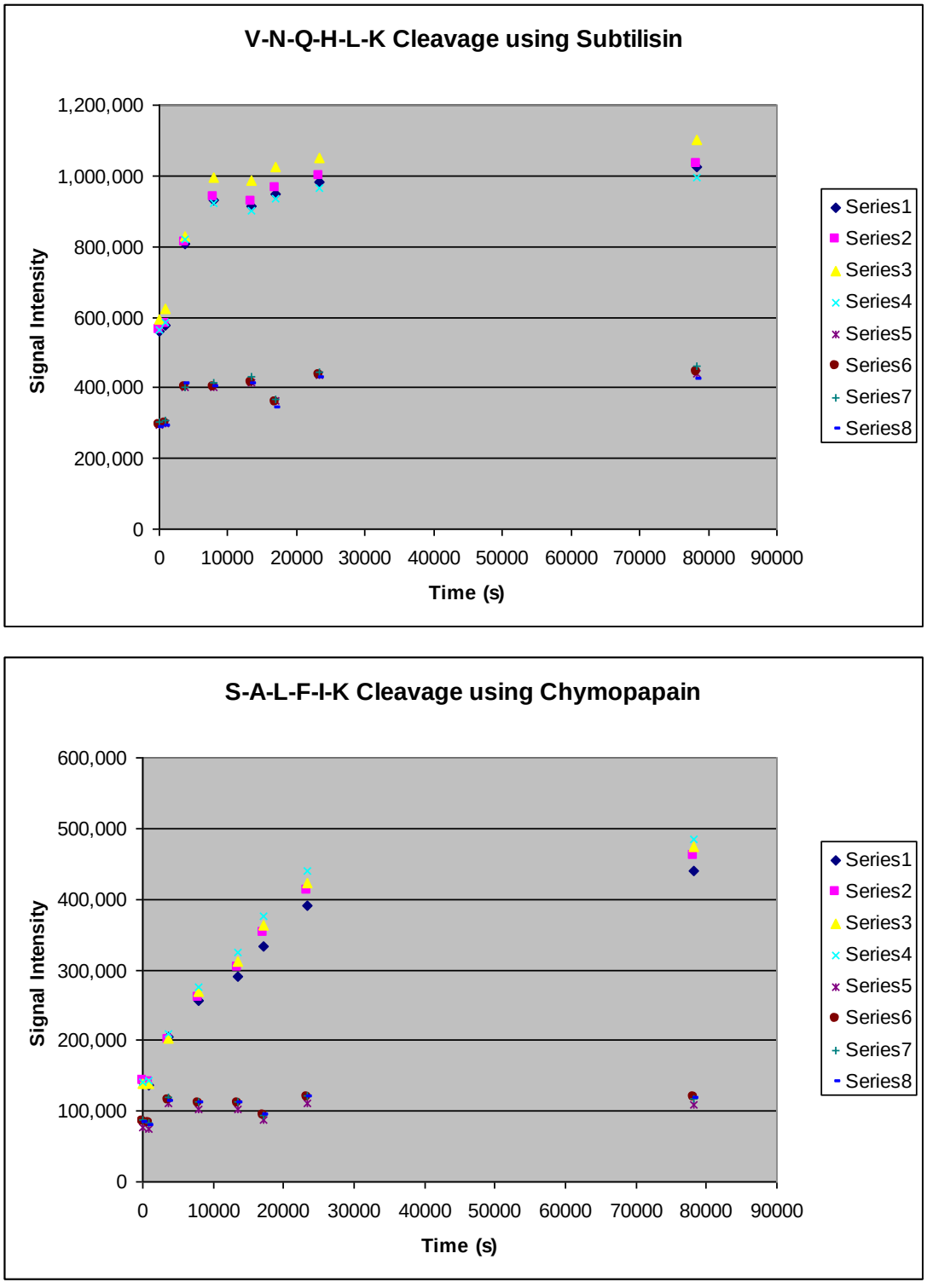


Figure 2.13 – Evaluation of V-N-Q-H-L-K with Subtilisin (top), and S-A-L-F-I-K with Chymopapain (bottom). Series from 1 to 4 represent the intensity in the fluorescein signal vs. time, in four different assays. Series 5 to 8 represent the intensity for Cy3 vs. time, in four different assays.

2.6. Results and discussion

In this investigation four different methodologies were assessed with regards to the determination of protease substrate specificity. For this a small library of 48 hexapeptides doubly labelled with FRET dyes were prepared and used with success. The assays using surface bonded constructs were not successful whether they were performed in 2D surfaces or on 3D surfaces. This allows us to conclude that the bonding of substrates limits the enzymatic activity in cases such as the ones explored in these assays. Not bonding the substrates to the surface and freeing them in a medium that is suitable for the assays, like glycerol droplets, microwells, or as in the final approach, in hydrophibic areas (using a in-house developed masked hydrophobic slide), opens the possibility of the use of all amino acid residues without any limitation, in the synthesis or in the assays (the constructs used were limited and no lysine and arginine had been used to avoid any bonding to the surface).

The best working assays for reactivity profiling were the ones carried out using the microwells and the hydrophibic platforms. The main advantage of these two applications resides in knowing the volume of enzyme solution printed and the complete "immobilization" of the substrates that was not observed in the enzyme spray assays reported in the literature.[25-27]

The assays described here can be carried out in a complete organized manner with a knowledge and control of the position of the substrates and enzymes and this allows the reactivities to be measured with high precision, at reduced cost (limited amounts of solvents, construction, and enzymes), in a very limited time frame (typically the library of constructs takes around 15 minutes to print and the enzyme under study less than 1 minute to print), and with a very reliable signal increase in fluorescence on hydrolysis (the fluorescein signal doubled in intensity). This investigation has also allowed us to identify substrates for all of the enzymes, with the most reported enzyme of the ones used (thermolysin), being the one with the most number of positive substrates identified (5 substrates; only Y-Q-L-E-N-K wasn't cleaved). For both subtilisin and chymopapain only one substrate was identified (V-N-Q-H-L-K and S-A-L-F-I-K respectively). In the case of

chymopapain a relative preference for lysine in the P_1 is known and this could partially explain the non-cleavage of the other peptide sequences used. Nevertheless, the cleavage of the substrate S-A-L-F-I-K and the preference of subsite S_2 for non-polar side chains might represent a different situation. Looking at **table 2.3** it is clear that if a preference for leucine for the subsite S_2 is the only criterion two other substrates should have been cleaved, namely A-L-F-I-A-K and S-L-F-I-A-K. There appears to be other parameters involved in chymopapain substrate specificity.

In the case of subtilisin, for which the sequences chosen have been reported in the literature[21-23] it was unexpected that only one of the constructs was cleaved. Nevertheless, for all the tested enzymes in this study substrates were identified which proves the reliability of the assays developed (**table 2.5**).

Table 2.5 – Identified substrates and controls from the presented in section 2.4.

Thermolysin	Chymopapain	Subtilisin	Controls
Y-G-G-F-L-K	S-A-L-F-I-K	V-N-Q-H-L-K	P-I-Q-Q-H-K
G-F-F-Y-Y-K			H-I-P-P-Y-K
E-A-L-Y-L-K			P-I-H-Y-Q-K
			H-P-I-Q-Q-K
Y-N-Q-F-F-K			I-H-P-Q-P-K
G-S-H-L-V-K			P-I-P-I-S-K

The design of the constructs worked, as well as the in-house platform. Also the control sequences which were not cleaved in the assays with the different enzymes. Therefore it is safe to say that a new methodology was successfully developed and optimized for the determination of protease substrate specificity.

Further studies could be carried out using the identified substrates for the screening of inhibitors of the enzymes, and also to extend the use of the methodology to identify new substrates for other proteases.

2.7. References

(1) Dahan, M.; Deniz, A. A.; Ha, T. J.; Chemla, D. S.; Schultz, P. G.; Weiss, S. *Chemical Physics* 1999, *247*, 85-106.

(2) Stryer, L. *Annual Review of Biochemistry* 1978, *47*, 819-846.

(3) LeBonniec, B. F.; Myles, T.; Johnson, T.; Knight, C. G.; Tapparelli, C.; Stone, S. R. *Biochemistry* 1996, *35*, 7114-7122.

(4) Sheppeck, J. E.; Kar, H.; Gosink, L.; Wheatley, J. B.; Gjerstad, E.; Loftus, S. M.; Zubiria, A. R.; Janc, J. W. *Bioorganic & Medicinal Chemistry Letters* 2000, *10*, 2639-2642.

(5) Marschutz, M. K.; Veronese, F. M.; Bernkop-Schnurch, A. *European Journal of Pharmaceutics and Biopharmaceutics* 2001, *52*, 137-144.

(6) Cuerrier, D.; Moldoveanu, T.; Davies, P. *Journal of Biological Chemistry* 2005, *280*, 40632-40641.

(7) Bycroft, B. W.; Chan, W. C.; Chhabra, S. R.; Hone, N. D. *Journal of the Chemical Society-Chemical Communications* 1993, 778-779.

(8) Bialy, L.; Diaz-Mochon, J.; Specker, E.; Keinicke, L.; Bradley, M. *TETRAHEDRON* 2005, *61*, 8295-8305.

(9) Diaz-Mochon, J.; Bialy, L.; Bradley, M. *ORGANIC LETTERS* 2004, *6*, 1127-1129.

(10) Metzker, M. L.; Lu, J.; Gibbs, R. A. *Science* 1996, *271*, 1420-1422.

(11) Merrifie.Rb; Stewart, J. M. *Nature* 1965, *207*, 522.

(12) Andrews, R. P. *Nature* 1986, *319*, 429-430.

(13) Atherton, E.; Brown, E.; Sheppard, R. C.; Rosevear, A. *Journal of the Chemical Society, Chemical Communications* 1981, 1151-1152.

(14) Sheppard, R. C.; Williams, B. J. *International Journal of Peptide and Protein Research* 1982, *20*, 451-454.

(15) Palasek, S.; Cox, Z.; Collins, J. *Journal of peptide Science* 2007, *13*, 143-148.

(16) Nguyen, T.; Francis, M. B. *Organic Letters* 2003, *5*, 3245 - 3248.

(17) Rawlings, N.; Barrett, A. *Biochemical Journal* 1993, *290*, 205-218.

(18) Rawlings, N.; Morton, F.; Kok, C.; Kong, J.; Barrett, A. *Nucleic Acids Research* 2008, *36*, D320-D325.

(19) Fricke, B.; Drossler, K.; Willhardt, I.; Schierhorn, A.; Menge, S.; Rucknagel, P. *Biochimica et Biophysica ACTA-Molecular Basis of Disease* 2001, *1537*, 132-146.

(20) Matsubar.H; Sasaki, R.; Singer, A.; Jukes, T. *Archives of Biochemistry and Biophysics* 1966, *115*, 324-325.

(21) Johansen, J.; Ottesen, M. *Comptes Rendus des Travaux du Laboratoire Carlsberg* 1968, *36*, 265-268.

(22) Kraus, E.; Kiltz, H.; Femfert, U. *Hoppe-Seylers Zeitschrift fur Physiologische Chemie* 1976, *357*, 233-237.

(23) Toogood, H. *Handbook of Proteolitic Enzymes*; 2 ed.; Elsevier: London, 2004.

(24) Diaz-Mochon, J.; Bialy, L.; Bradley, M. *Chemical Communications* 2006, 3984-3986.

(25) Gosalia, D.; Diamond, S. *Proceedings of the National Academy of Sciences of the United States of America* 2003, *100*, 8721-8726.

(26) Gosalia, D.; Salisbury, C.; Ellman, J.; Diamond, S. *Molecular & Cellular Proteomics* 2005, *4*, 626-636.

(27) Gosalia, D.; Salisbury, C.; Maly, D.; Ellman, J.; Diamond, S. *Proteomics* 2005, *5*, 1292-1298.

(28) Gosalia, D. N.; Salisbury, C. M.; Ellman, J. A.; Diamond, S. L. *Molecular & Cellular Proteomics* 2005, *4*, 626-636.

(29) Khan, F.; Zhang, R.; Unciti-Broceta, A.; Diaz-Mochon, J. J.; Bradley, M. *Advanced Materials* 2007, *19*, 3524-3529.

Chapter Three

Every long walk starts with a single step …and then you want to run!

The purpose of this chapter is to report on some of the different techniques used over recent years in the synthesis of "chemical diversity", namely microwave heating and fluorous chemistry. The use of these two techniques regarding the preparation and application of a traceless linker are reported in this chapter.

3. "Speeding" Up Chemistry: Microwave and Fluorous Phase Techniques

The difference in concept between drug discovery today and 20 years ago is underlined by the difference in number of compounds produced.[1] If the focus used to be on pushing a small number of rationally designed drugs through preclinical and clinical studies, there is today a clear trend and willingness to improve the chances of discovering a successful drug by one of the simplest methods available: increasing the repertoire of synthesized compounds. There have been a number of practical solutions on how to address this issue of enhancing the output of unique chemical identities e.g. combinatorial and parallel synthesis.

Even though many of these techniques in themselves are modern, there has been an interest in improving these methods, and inventing new techniques, to further increase the speed of synthesis. These efforts can be divided into techniques that increase either the speed of *synthesis*, such as microwave techniques, or the speed of purification[2], such as fluorous techniques.

Microwave chemistry is today an important and accepted method of heating in chemistry.[3-5] The attractions of microwave-heated technologies are many. For example, reaction times can in numerous cases be shortened from hours or days to minutes or seconds, reactions can be executed safely, with pressure- and temperature

control, in an interlaboratory reproducible manner. Fluorous chemistry and separation, which today is a general term for several interrelated subcategories of chemistry using the unique properties of compounds with high fluorine content, is an area where there have been invigorating developments.[6-9] Even though the immediate gain of fluorous chemistry is in the separation steps, it could be regarded as an alternative in synthesis with advantages that cover all the steps from the beginning of a reaction to the purification of the final product.

3.1. Microwave enhanced chemistry

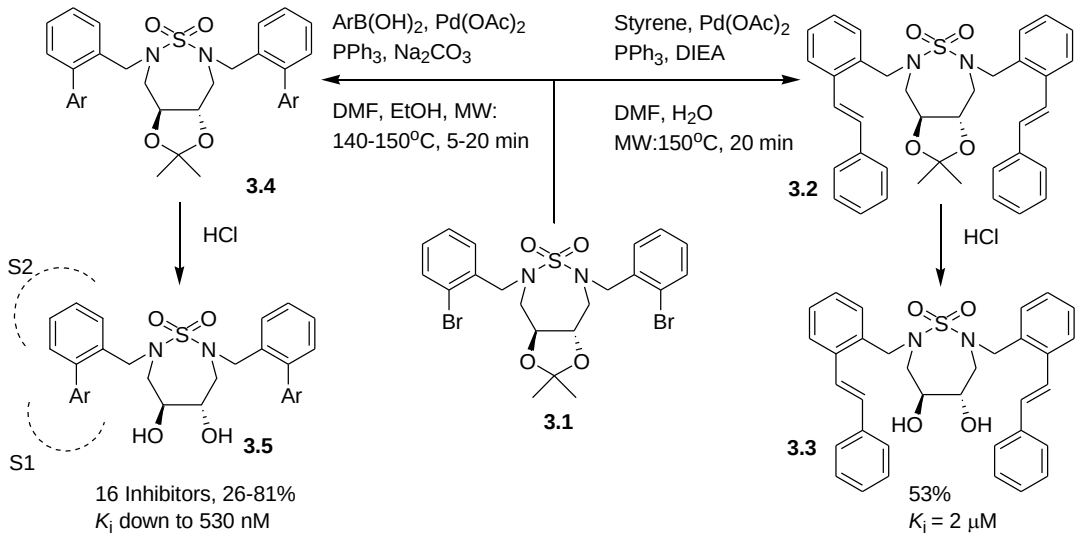
Although the use of microwave technology in drug-discovery started earlier[10-15], the great push occurred in the 1990's with the relevant contribution of Hallberg[16] who early on emphasized the potential of combining microwave heating and transition metal catalysis in the development of sophisticated lead molecules. Exploiting high-density microwave heating, Hallberg's[17] research group demonstrated that it was possible to drive palladium-catalysed transformations to completion in minutes instead of hours, and that new HIV-inhibitors could be rapidly generated.[18,19] When the benefits of combining microwave heating with solid-phase chemistry and polymer-supported reagents were recognized, these concepts were investigated in detail by Kappe[20] and Ley[21]. Convincing evidence that dedicated microwave equipment indeed accelerated both reaction development and compound-production accumulation was obtained and today, microwave reactors are an integrate part of the standard instrumentation in most modern chemistry laboratories.

Since a large number of recent reviews on microwave-assisted organic and combinatorial chemistry are available[4,22], no introduction to the theory behind microwave heating will be given.[23] I will only conclude that high-density microwave irradiation produces efficient internal heating of polar liquid systems, resulting in an even and rapid heating throughout the sample. Importantly, almost all organic reaction systems can be accelerated utilizing this energy source. The different types of commercial reactors available will not be discussed, but typically reactions are conducted in 0.5-5.0 mL, septum-sealed reaction vessels employing temperature-controlled single-mode synthesizers.[24]

Many applications of microwave chemistry have involved small-scale synthesis of peptidomimetic or transition state-mimicking core structures using palladium-catalysed coupling reactions.[18,19,25,26] The success of this strategy is based on the high selectivity associated with transition-metal catalysis together with the simplicity and short reaction times achieved with modern single-mode microwave reactors. Furthermore, to allow fine-tuning of the molecular composition of a lead scaffold with the ultimate goal of obtaining improved biological activities, solubility, metabolic stability, and pharmacokinetic properties, large sets of modified analogues are easily obtained via microwave assisted scaffold decoration.

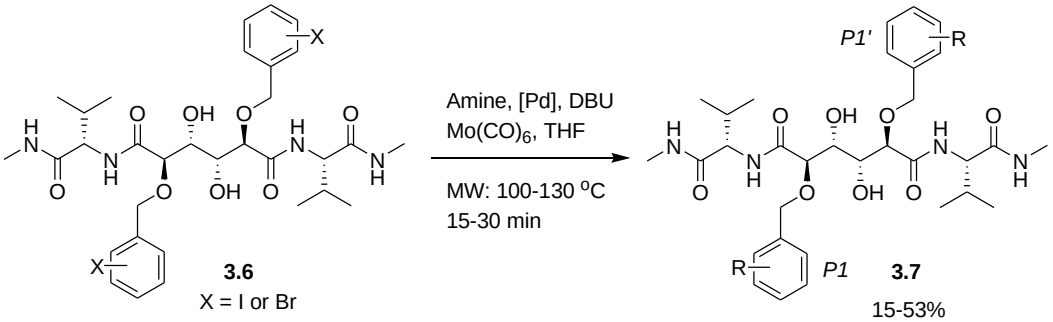
3.2. Examples: Applications in Medicinal Chemistry

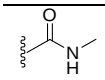
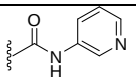
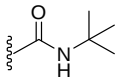
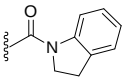
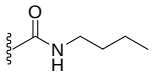
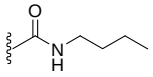
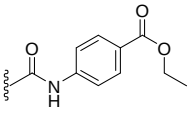
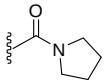
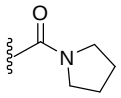
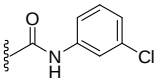
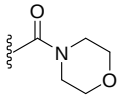
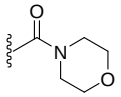
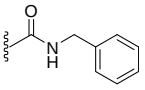
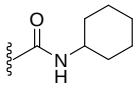
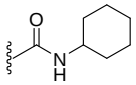
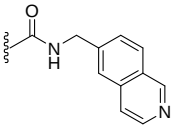
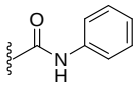
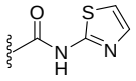
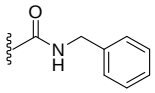
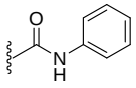
As a result of the emergence of drug resistant in HIV-1 strains and severe side effects of current anti-HIV drugs, continued interest in HIV-1 protease inhibitor research is still highly motivated. In this field, the activity of tetra-substituted cyclic sulfamide protease inhibitors incorporating the central dihydroxyethylene transition state isostere, has been well documented.[26,27] A set of microwave enhanced Suzuki and Heck couplings executed by 5-20 minutes irradiations were used to elongate the P_2/P_2' side chains of the cyclic HIV-1 protease inhibitor template as presented in **scheme 3.1**. The products after deprotection were screened for HIV-1 protease activity, identifying a relatively potent dibenzofurane inhibitor that possessed a K_i of 530 nM[28]. Thus, in the attempt to develop a new class of HIV-1 protease inhibitors, the benzylic P_2/P_2' groups were substituted in the *ortho*-position reach the S_1/S_1' pockets.[28]



Scheme 3.1: Synthesis of dibenzofurane based HIV-protease inhibitors.[28]

Another class of HIV-1 protease inhibitors are the linear 1,2-dihydroxyethylene-based compounds.[29] In a project searching for novel binding modes to the protease, the 2-iodobenzyloxy and 3-bromobenzyloxy P_1/P_1'–substituted structures depicted in **scheme 3.2** were used as arylpalladium precursors.[30] Two series of microwave-promoted aminocarbonylations were performed in which $Mo(CO)_6$ served as a convenient solid source of carbon monoxide.[17,31] The discovery of active single-digit nanomolar inhibitors containing large phenyl amide *ortho*-substituents in the P_1/P_1' positions indicates that larger groups might be tolerated in this part of the S_1/S_1'-pocket than previously believed.[30]

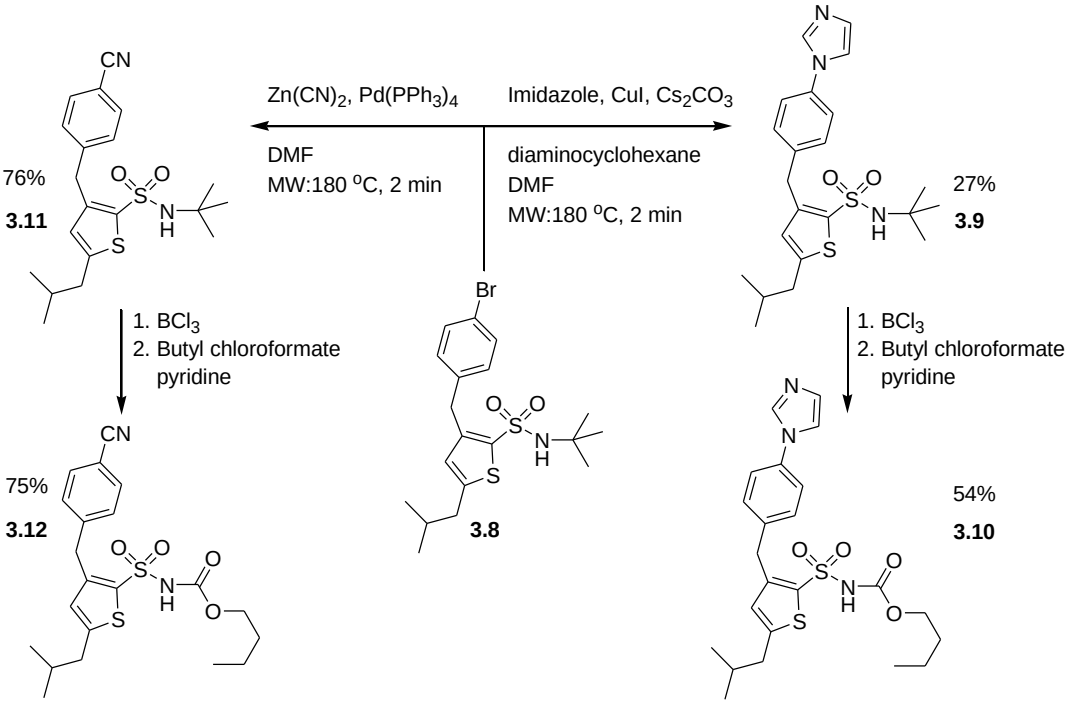


ortho-I, R	K_i (nM)	*ortho*-I, R	K_i (nM)	*meta*-Br, R	K_i (nM)
	>5000		70		20
	700		800		3
	600		200		200
	400		6		300
	>5000		2000		20
	200		800		7
	700				2
	7		900		

Scheme 3.2: Microwave-promoted aminocarbonylations using $Mo(CO)_6$, and K_i values of the synthesized substrates (reproduced with permission).[17,31]

The human octapeptide angiotensin II mediates its effects at two major receptors, named AT_1 and AT_2. While the much-studied AT_1 receptor is involved in the control of blood pressure, electrolyte, and fluid balance, the function of the AT_2 receptor has been more difficult to elucidate. Recent studies suggest that activation of the AT_2 receptor, among other effects, stimulates alkaline secretion by the duodenal mucosa in rats.[19] Based on this information, Alterman[18] reported the discovery of

selective AT_2 receptor agonists using a number of microwave-mediated transformations. A small part of this extensive project is shown in **scheme 3.3**, where two compounds **3.10** and **3.12** of modest activity were prepared by microwave heated coupling reactions.[18]



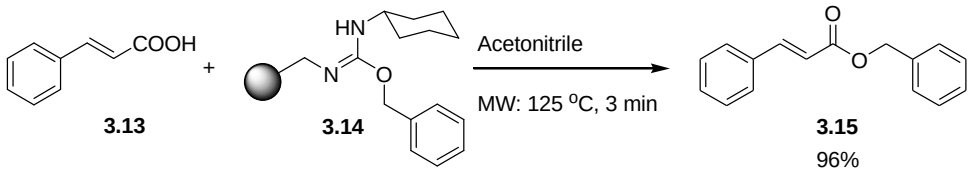
Scheme 3.3: Microwave-assisted Ullman-type reaction and palladium catalysed nitrile coupling used in the synthesis of AT_2 receptor agonists.[18]

In the top left case, **3.11** was prepared through a palladium-catalysed nitrile coupling, rapidly executed by 2 min microwave irradiation, while the **3.9** phenyl-imidazole moiety was created using a very sluggish Ullman-type reaction.

3.3. Applications in Solid-Phase Chemistry

Microwave chemistry is beginning to play a greater role in solid-phase chemistry and in cases where polymer-supported reagents are utilized.[3,22] Drawbacks of these methodologies include slow reaction kinetics. An impressive example of a microwave accelerated and polymer assisted process was described by Linclau[32]. In

this report, a set of carboxylic acids were fully *O*-alkylated after only 3-5 min of microwave heating (**scheme 3.4**).



3.13 **3.14** **3.15**
96%

Scheme 3.4: Microwave accelerated *O*-alkylation of carboxylic acids in solid support.[32]

3.4. Fluorous Applications in High-Throughput Chemistry

A fluorous phase Pummerer reaction was introduced by Procter *et al.*.[33] This report was based on the need to address problems associated with difficulties in optimization and monitoring of solid-phase processes. Good to excellent yields of heterocyclic scaffolds, mainly oxindoles, were reported after the development of a high-throughput fluorous phase synthesis strategy.[33]

A somewhat different approach was used by Vincent *et al.*[34], when in phase switching reactions, pyridyl-labelled substrates and products were used. The pyridyl-coating tag can be thought of as a masked phase tag, which allows for phase switching with the help of a heavy fluorous copper(II)-carboxylate complex. Comparison with a non-fluorous system indicated that the release of the strongly coordinating pyridine linker was avoided in the fluorous approach.[34]

A fluorous catch and release method was used by Zhang's group in the synthesis of disubstituted pyrimidines.[9] A fluorous route to hydantoins and thiohydantoins was also reported by the same group.[9]

3.5. Microwave–Enhanced Fluorous Chemistry

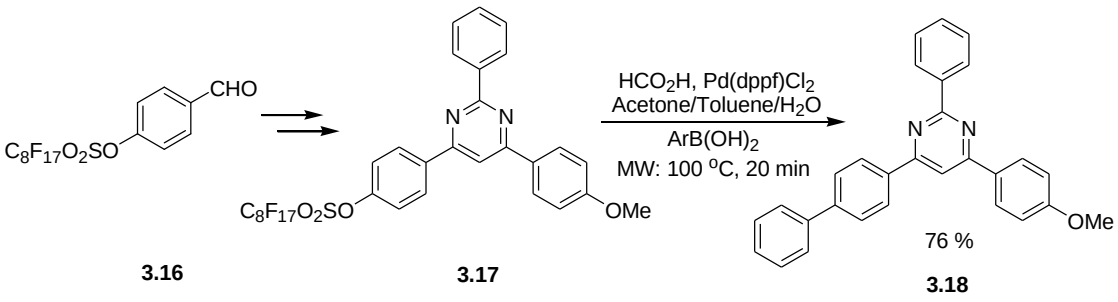
The combination of microwave heating and fluorous chemistry is of considerable interest, as this approach combines fast chemistry and easy separation.

A fluorous palladium pincer complex for use in Heck reactions was reported by Curran.[35] The tridentate ligand ensured stability under the high temperatures (140^{o}C) needed for the reaction to go to completion. The complex was recovered and reused three times without any sign of lowered catalytic activity.[7]

An *in situ* carbonylation procedure was shown by Larhed to be compatible with recycling of the fluorous catalyst used in the reaction. The catalyst was collected five times by a two-phase liquid fluorous extraction, with yields varying only slightly between experiments.[36]

Zhang has explored the possibilities of using more highly fluorous versions of the triflate group in palladium catalysed reactions. The $C_8F_{17}O_2SO$- leaving group worked efficiently with organoboronic acids in Suzuki couplings[37], with the higher degree of fluoricity enabling purification by F-SPE. The authors also underlined the higher speed of the reaction under microwaves as compared to classic heating as well as the ease of separation connected with the use of fluorous tags.

Microwave-assisted fluorous Ugi reactions were presented where the reaction times and convenient separation techniques appeared more attractive than the corresponding room temperature methods with traditional scavenging techniques.[37]



Scheme 3.5: Suzuki coupling using $C_8F_{17}O_2SO$- as a leaving group.[37]

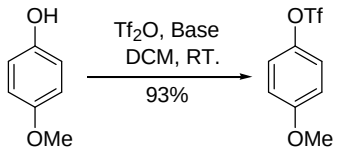
A parallel synthesis method for the synthesis of a dihydropteridine library in five steps was recently published by Zhang.[38] The synthetic route was based on the use of fluorous amino acids, with the final cyclization prompted by use of microwave heating.

3.6. Synthetic transformations of aryl triflates and sulfonates

3.6.1. Formation of aryl sulfonates

The increasing use of vinyl and aryl trifluoromethanosulfonates (triflates) is due to their facile preparation from carbonyl compounds and phenols and to the excellent leaving group properties of the trifluoromethanesulfonate group. The leaving group ability of aryl triflates relative to aryl halides may be generalized in the following order I > Br ≈ OTf >> Cl, although this is obviously to some extent reaction dependent. Recently, the many synthetic transformations of vinyl and aryl triflates were reviewed by Ritter.[39]

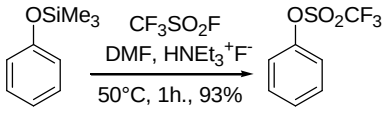
Aryl triflates can be prepared from phenolic substrates by way of a triflating reagent in the presence of a non-nucleophilic base (e.g. triethylamine or DIPEA), (**scheme 3.6**).[40]



3.19 **3.20**

Scheme 3.6: Formation of an aryl triflate from 4-methoxyphenol.[40]

They can also be prepared by fluoride ion catalysis of perfluoroalkylsulfonyl fluorides with aryl silyl ethers (**scheme 3.7**).[41] This method commonly used for the preparation of triflates and nonaflates, facilitates the use of nonafluorobutanesulfonyl fluoride, a relatively stable liquid, easily handled under regular laboratory conditions.

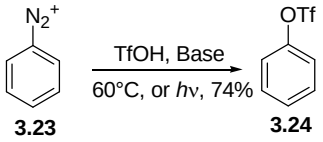


3.21 **3.22**

Scheme 3.7: Formation of perfluoroarylsulfonates from trimethylsilyl ethers.[41]

The preparation of aryl triflates by thermal or photochemical decomposition of the intermediate aryl diazonium salt is also possible (**scheme 3.8**).[42] Given that the

formation of the diazonium species from an aniline precursor is a reaction tolerant of many other nucleophilic functions, this method is suitable for the preparation of hydroxyphenyl triflates, which would be difficult to prepare by other methods.[42]

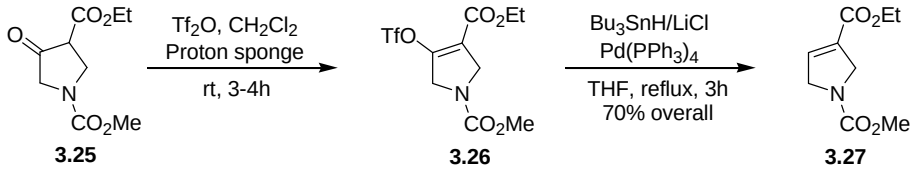


Scheme 3.8: Preparation of aryl triflates from aryl diazonium salts.[42]

The perfluoroalkylsulfonate group allows electrophilic aromatic substitution reactions such as nitration, sulfonation and halogenation to be performed in the *ortho* and *para* position of the aryl substrates bearing such directing sulfonate groups.

3.6.2. Transfer hydrogenation of aryl triflates

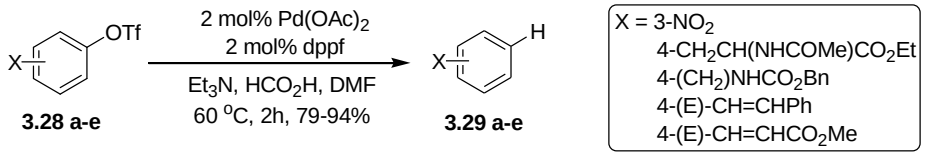
In the presence of a catalytic quantity of Pd (0), and a suitable hydride donor source, aromatic triflates are reduced cleanly, through a process known as transfer hydrogenation. Several different hydride sources have been used with good efficiency, including both tributyltin hydride and triethylsilane. In 1991, Dupré *et al.*[43] reported the 2-step reduction of a pyrrolidone **3.25** to the unsaturated cyclic ester **3.27** via the triflate intermediate **3.26** (**scheme 3.9**).[43]



Scheme 3.9: Transfer hydrogenation of a vinyl triflate.[43]

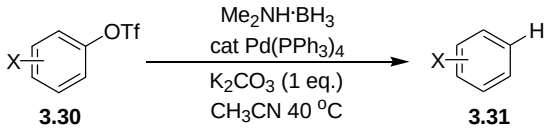
Triethylammonium formate, formed by the *in situ* reaction of formic acid with triethylamine has also been used as an exceptionally mild and selective hydride source in transfer hydrogenation reactions of triflates, in the presence of catalytic Pd (0) and a suitable ligand and a range of other potentially reactive functionalities.[44]

Sterically hindered or electron-rich phenyl triflates require the use of chelating bis-phosphine ligands such as dppf[45] or dppp[46] (**scheme 3.10**).



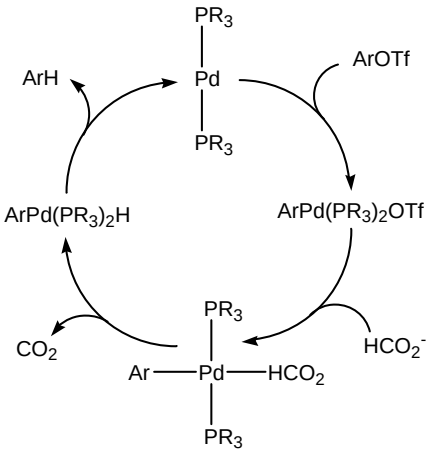
Scheme 3.10: Transfer hydrogenation of a aryl triflates.[45,46]

Lipshutz *et al.*[47] demonstrated reductions of arylperfluorosulfonates **3.30** using dimethylamine·borane complex **3.31** ($Me_2NH{\cdot}BH_3$) also catalysed by Pd(0).[47] Using these conditions, reductive cleavage was generally effected in 3.5 hours at only 40 ºC. This combination of reagents also represents an extremely mild method for achieving reductive cleavage of activated sulfonates (**scheme 3.11**).



Scheme 3.11: Reductive cleavage of aryl triflates using $Me_2NH{\cdot}BH_3$.[47]

Ortar *et al.*[48] have postulated that the mechanism of transfer hydrogenation of triflates involves the oxidative addition across the triflate carbon-oxygen bond to a Pd(0) phosphine ligand. Displacement of the triflate by a formate ion, followed by loss of carbon dioxide results in the formation of an arylpalladium (II) hydride species. Reductive elimination of this complex regenerates the active palladium catalyst, and releases the reduced aromatic compound from the catalytic cycle (**scheme 3.12**).[48]



Scheme 3.12: Palladium-catalyzed transfer hydrogenation of an aryl triflate.[48]

If deuterated formic acid is used in place of the regular protic acid, then in conjunction with triethylamine, a deuterated product is produced, and this methodology has been expanded to allow isotopic labelling of triflate precursors[48]. Isotopic labelling, for example using deuterium, is an important technique used in the elucidation of reaction mechanisms and kinetic rate calculations, and hence the ability to selectively label a molecule is of high research value. Since hydrogen and deuterium are isotopes of the same element, the chemistry of the carbon-hydrogen and carbon-deuterium bond will be identical. However, since these are atoms of different mass, the vibrational frequencies and hence bond dissociation energies will differ, resulting in a difference in relative bond strength. Therefore the rate of reaction about these bonds will be significantly different, allowing for kinetic evaluation.

3.6.3. Metal catalyzed cross-coupling of sulfonates and triflates

Although aryl and vinyl triflates have recently received attention regarding participation in metal-catalysed cross-coupling reactions, they have traditionally been considered too unreactive, and most reported examples employ aryl and vinyl bromides and iodides. Indeed, when triflates are used directly in place of bromides or iodides, without expecting the need for considerable optimization of conditions, disappointment usually prevails. However, even under traditional cross-coupling

conditions the use of aryl triflates will most often result in the formation of some of the desired product, and this seminal finding paved the way for another generation of cross-coupling methodology. 1995 saw the first of many publications to bring forward unactivated aryl and alkyl sulfonates into the cross-coupling spotlight, and today this methodology is continually advancing. Percec addressed the shortfalls of the aryltriflate and arylmesilate, and developed a new cross-coupling protocol for reaction with arylboronic acids.[49]

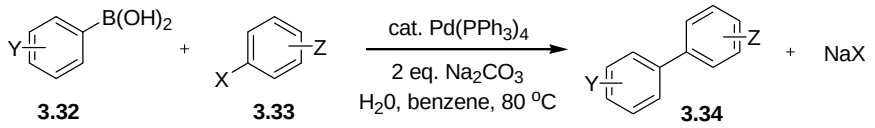
Sulfonates are base sensitive and prone to hydrolysis under aqueous conditions, however Percec *et al.*[49] found that the use of the weak base K_3PO_4 when finely powdered and suspended in polar solvents such as dioxane or THF fulfil these non-aqueous requirements. Also recognized was the need for a palladium/ligand system more active than $Pd(PPh_3)_4$ and the dinuclear catalyst $PdCl_2$(dppf) was demonstrated as a good successor. Most notably Percec *et al.* saw the need for addition of lithium chloride to the reaction mixture, a common additive used in the Stille cross-coupling reactions between triflates and stannates.[50] Alkali metal halides are expected to prevent premature catalyst decomposition at the oxidative addition stage, the crucial first step of the catalytic cycle. The product of oxidative addition of palladium into the arene sulfonate carbon-oxigen bond has been proposed as being "catalitically incompetent", although ligand substitution with chloride ions results in a catalytically active species.[51] Hence lithium chloride rescues the reaction from stalling at stage of conception of the catalytic cycle. Using the modified reaction conditions outlined above, Percec *et al.*[49] demonstrated the successful cross-coupling of aryl triflates in place of the more commonly employed bromide and iodide electrophiles in comparable yield, a significant achievement in itself. However, a still more impressive advancement brought various unactivated aryl sulfonates onto the list of coupling contenders. Realising the unreactivity of aryl sulfonates towards Pd(0) species, Percec used instead the analogous Ni(0) catalyst. Previously, Percec *et al.*[52] had observed the rapid oxidative addition of aryl sulfonates to Ni(0) complexes as a result of the increased nucleophilicity of Ni(0) over Pd(0).[52] The resulting Ni(II) complexes were found to undergo symmetrical homocoupling in the absence of other electrophilic species to give biaryl products in high yield. Thus when phenylboronic acid and K_3PO_4 suspended in THF was added to the reaction mixture, as expected,

successful Suzuki-type cross-coupling was observed in moderate yield. This is a significant finding and represents the first example of Ni(0) catalysed coupling between boronic acids and aryl sulfonates other than triflates. Not only is there financial saving in terms of the catalyst, but also these reaction conditions produce similar yields of coupled compounds. However, there are two notable drawbacks to be considered when using Ni(0) catalysis in this way. Primarily, Ni(0) species are deactivated by protic sources, thus reactions must be conducted under strictly anhydrous conditions. Also, *in situ* reduction of Ni(II) to Ni(0) does not occur spontaneously in the presence of a base as with Pd(II) sources like $Pd(OAc)_2$ and requires zinc dust. Whilst these constrains do not detract significantly from the virtues of Ni(0) catalysis, they do prompt the experimentalist the use of aryl triflates which are still reactive enough to undergo Pd(0) catalysed cross-coupling with boronic acids without the need for special reaction conditions.

3.6.4. The Suzuki-Miyaura cross-coupling reaction

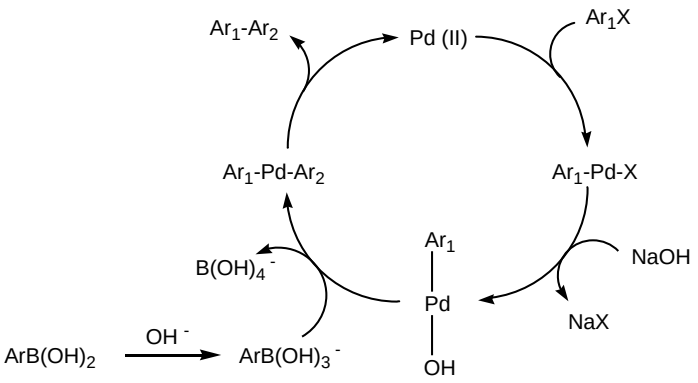
Interest in the use of aryl triflates was stimulated greatly when Suzuki reported a palladium-catalysed cross-coupling reaction with aryl boronate esters and boronic acids in the presence of a base. In contrast to preceding methods of aryl carbon-carbon bond formation, the extremely mild conditions and high functional group tolerance of the Suzuki-Miyaura reaction quickly established this as one of the most powerful methods for biaryl formation.[53] Given the significance of the biaryl motif, both from a pharmaceutical and total synthesis perspective, this new ease with which biaryl compounds may be prepared has produced an enormous interest in this methodology. It is then not surprising that today the literature is awash with reports of palladium-catalysed transformations performed upon unsaturated halides and triflates and resulting in the creation of new bonds at sp^2 carbon centers.

During Suzuki's preliminary cross-coupling reactions[54] conducted in benzene, he found that in addition to a catalytic quantity of $Pd(PPh_3)_4$, 2 equivalents of aqueous Na_2CO_3 are required for successful coupling of substituted phenyl boronic acids **3.32**, and aryl halides **3.33** (**scheme 3.13**).[54]



Scheme 3.13: The Suzuki-Miyaura palladium-catalysed cross-coupling.[54]

The Suzuki-Miyaura reaction is very broad with regard to optimal conditions, and it is usually not possible to state a heavily generalized procedure for anything other than the simplest reactants. Considerations of literature examples reveal that a suitable Pd(0) catalyst depends both on the electronic and steric properties of the boronic acid and aryl halide or triflate coupling partners. $Pd(PPh_3)_4$ was the original catalyst system used by Suzuki[54] and continues to be the most widespread catalyst employed in the Suzuki-Miyaura reaction today. However, many other Pd(0) ligand systems have been exploited with advantage in more exotic applications. Although by no means an exhaustive listing, the following catalytic systems have proven highly efficient: $PdCl_2(PPh_3)_4$[55-57], $PdCl_2(dppf)$[58], $Pd(OAc)_2/PPh_3$[59], $Pd(OAC)_2$/dppp[60], $Pd(dba)_3$[61], $Pd(OAc)_2/P(o\text{-}Tol)_3$[62,63]. In addition, a very wide range of solvents/base mixtures has been employed, each having been optimized for an individual case. The most commonly utilized solvents in order of increasing polarity are PhMe, DME, THF, EtOH, and DMF. However, the selection of solvent is to some extent dictated by the base required. Bases which are commonly used include the group I carbonates Na_2CO_3[54], K_2CO_3[64], Cs_2CO_3[65], and also $Ba(OH)_2$[66], and K_3PO_4[49]. Gronowitz *et al.*[67,68] reported the use of DME and aqueous Na_2CO_3 as a solvent base system ideal for minimizing deboronation commonly observed as a side reaction of electron-rich boronic acids.

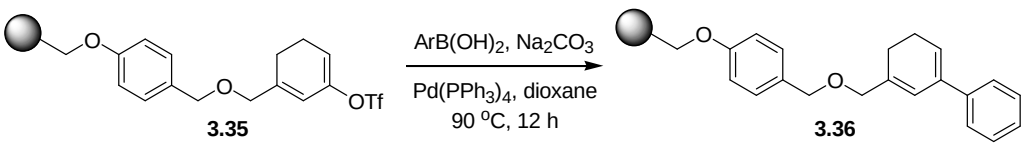


Scheme 3.14: Catalytic cycle of the Suzuki-Miyaura cross-coupling reaction.[54]

Providing the reaction in conducted under an inert atmosphere, the Suzuki-Miyaura catalytic cycle (**scheme 3.14**) remains active until one of the starting materials is consumed. Unlike many other organometallic reactions, the Suzuki-Miyaura reaction may be conducted in the presence of water, and indeed in most cases when an inorganic base is used, water is added in small amounts to allow at least partial dissolution of the base.

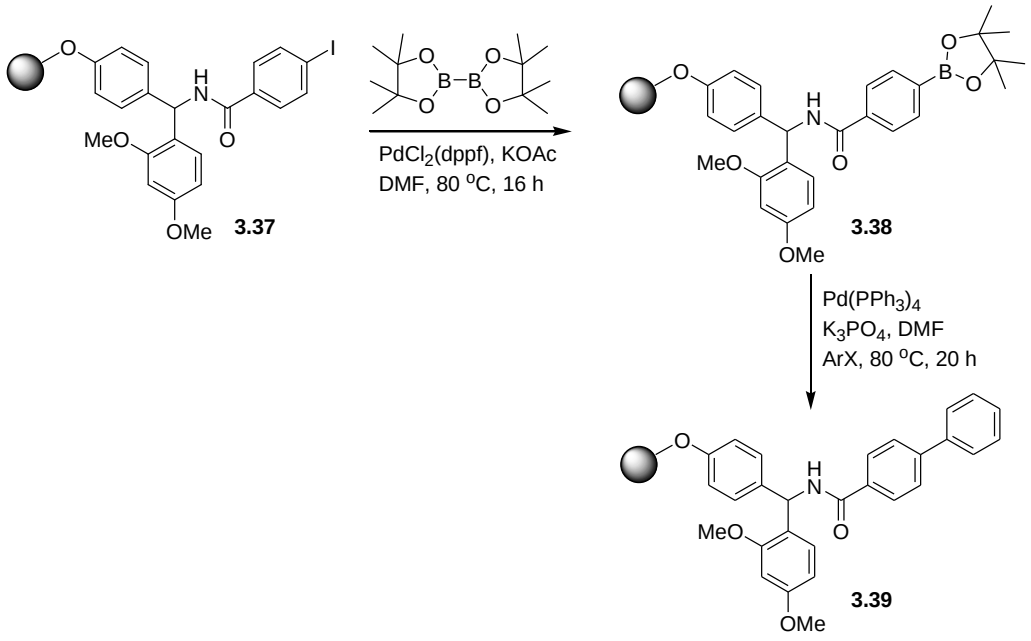
3.6.5. The Suzuki-Miyaura cross-coupling reaction on solid phase

The efficiency of the Suzuki-Miyaura reaction has translated very well to solid-phase conditions and there are a great number of high yielding applications to complement the analogous solution-phase reactions.[69-74] Fraley *et al.*[75] demonstrated the cross-coupling reaction of an immobilised vinyl triflate **3.35**, with several arylboronic acids during a synthesis of resin bound bicyclic compounds **3.36** (**scheme 3.15**).



Scheme 3.15: Suzuki-Miyaura cross-couppling of an immobilised vinyl triflate.[75]

Piettre *et al.*[76] demonstrated a dual-stage Suzuki-Miyaura cross-coupling reaction of dipinacolatoborane with resin bound aryl iodides **3.37** in the synthesis of immobilised boronate esters **3.38**, prior to reaction with aryl halides to give supported biaryl derivatives **3.39** (**scheme 3.16**).[76] In the presence of catalytic Pd(0), the combination of DMF and K_3PO_4 as solvent and base respectively, was found to give the best coupling yield.



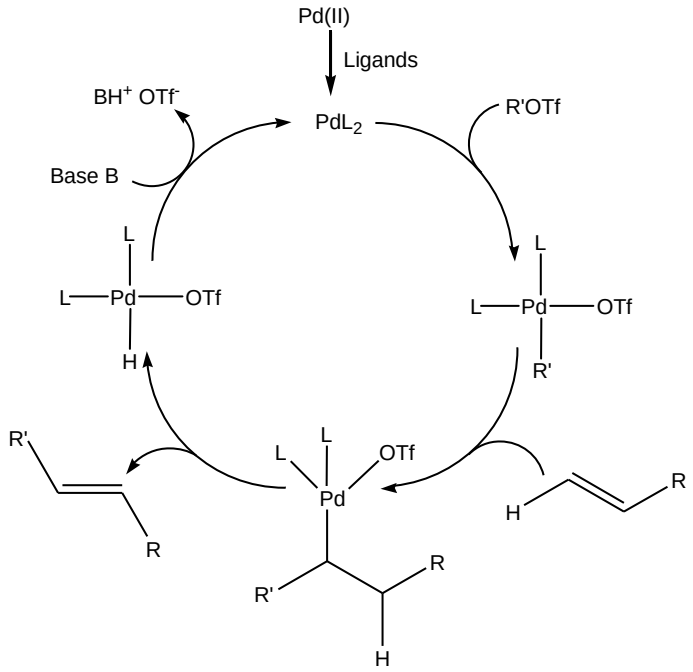
Scheme 3.16: Suzuki-Miyaura cross-coupling of a resin bound aryl iodide.[76]

However, the hydrophobic nature of cross-linked polystyrene resins regularly used as scaffolds for solid-phase synthesis, often requires the use of slightly modified reaction conditions. Inorganic bases used in solution-phase reactions often do not provide a homogeneous system in solvents compatible with polystyrene based resins, and weak organic bases such as Et_3N in DMF have been used instead with very good results.[63]

3.6.6. The Heck cross-coupling reaction

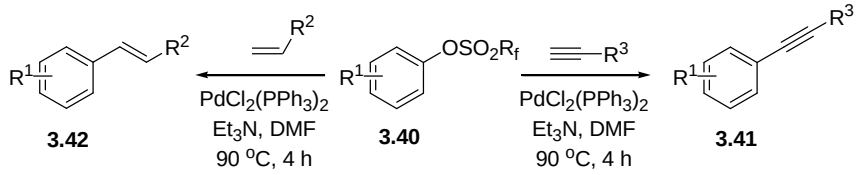
The Heck cross-coupling reaction between aryl or vinyl halides or sulfonates, and alkenes represents another palladium-catalysed process which has developed into

an extremely powerful tool for the construction of sp^2 carbon-carbon bonds.[77] In common with the Suzuki-Miyaura cross-coupling, the first step of the Heck reaction (**scheme 3.17**) is the oxidative addition of a palladium (0) species with an aryl or vinyl triflate (or bromide or iodide). Insertion of an alkene into this 16 electron species results in carbopalladation; rapid β-hydride elimination then results in the formation of a disubstituted alkene. The catalytic 14 electron Pd(0) species is the regenerated by reaction of the eliminated Pd(II) entity with a base, and the cycle continues until one of the substrates is exhausted.



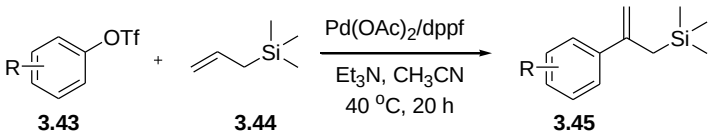
Scheme 3.17: The Pd(0) catalysed cycle of the Heck cross-coupling reaction.[77]

The first report of a Heck cross-coupling reaction with arylfluoroalkylsulfonates **3.40**, was made by Chen *et al.*.[78] Optimized reaction conditions revealed $Pd(PPh_3)_2Cl_2$ in combination with triethylamine produced very good yields of substituted styrenes **3.42**, and phenyl acetylenes **3.41** (**scheme 3.18**).[78]



Scheme 3.18: Heck reaction of aryl triflates with olefins and acetylenes.[78]

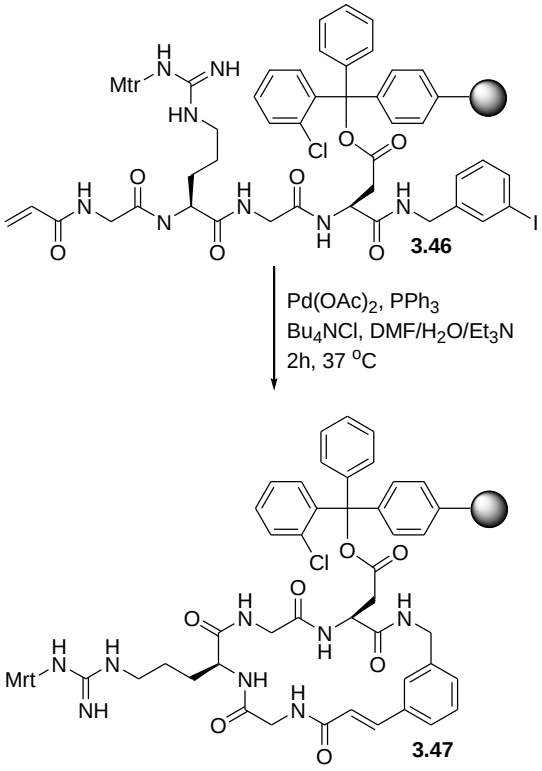
Recently, Olofsson *et al.*[79] reported the solution-phase Heck reaction of aryl triflates **3.43** with allyltrimethylsilane **3.44** using $Pd(OAc)_2$ in combination with dppf, resulting in the highly regioselective formation of branched β-products **3.45** in good yields (**scheme 3.19**).[79]



Scheme 3.19: Heck coupling of an aryl triflate with allylmethylsilane.[79]

3.6.7. The Heck cross-coupling reaction on solid-phase

Just as the Suzuki-Miyaura cross-coupling reaction has lead to good results on the solid-phase, so too has the Heck reaction, again increasing the repertoire of useful metal-catalysed bond forming processes available to the solid-phase chemist. At this time there exists an enormous number of publications citing reference to the Heck reaction on solid-phase.[80-82] The intramolecular Heck reaction is a very useful method for the formation of 5, 6 and 7 membered rings fused to aromatic nuclei. Akaji *et al.*[83] successfully demonstrated the use of the Heck cross-coupling reaction to perform an intramolecular macrocyclisation upon a bifunctionalised peptidic residue bearing both an aryl iodide and acrylate function.[83] The linear precursor **3.46**, immobilised on chlorotrityl chloride substituted polystyrene was cyclised in the presence of Pd(0) to yield a 20-membered macrocycle **3.47**, released from the support upon acid mediated cleavage (**scheme 3.20**).[83] Bu_4NCl is used as a phase-transfer agent, permitting the use of water with the supported substrate.

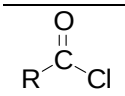
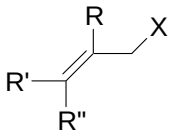
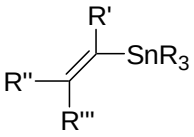
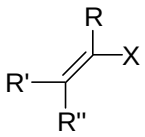
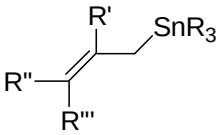
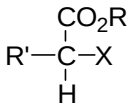


Scheme 3.20: Macrocyclisation using the Heck reaction.[83]

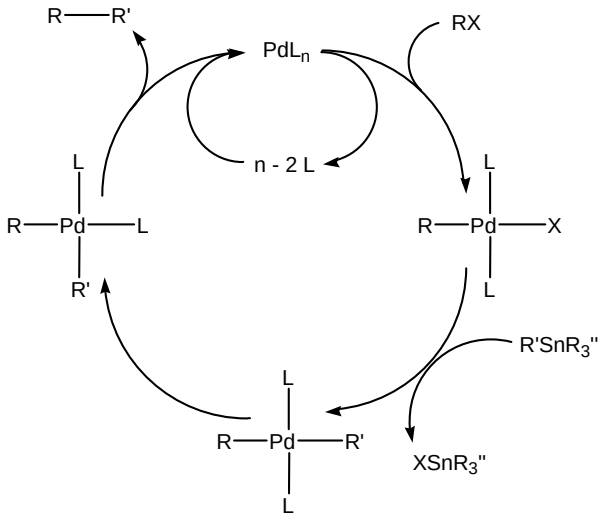
3.6.8. Stille cross-coupling.

The Stille cross-coupling reaction involves the coupling of organotin reagents with electrophiles (see **table 3.2**) catalyzed by palladium under mild conditions in high yields. A wide variety of organic halides can be coupled, either directly (**scheme 3.21**) or in the presence of carbon monoxide (to yield a ketone, RCOR').[50] Moreover, a wide variety of functional groups (*e.g.*, CO_2R, CN, OH, and CHO) are tolerated on either partner. Usually retention of configuration on the double bond is observed regardless of which reactant contains the double bond.

Table 3.2. – Tin-based reagents and electrophilic compounds suitable for Stille coupling reactions.[50]

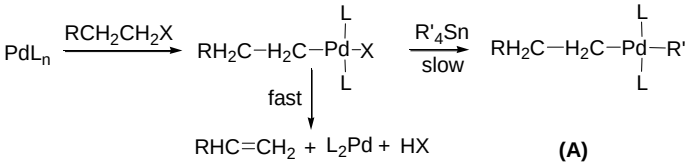
Electrophiles		Tin-based reagents
		$H-SnR_3$
	[X = Cl, Br]	$R'C\equiv C-SnR_3$
$ArylCH_2-X$	[X = Cl, Br]	
	[X = I, OTf]	$Aryl-SnR_3$
$Aryl-X$	[X = Br, I]	
	[X = Br, I]	$ArylCH_2-SnR_3$ $H_{2n+1}C_nSnR_3$

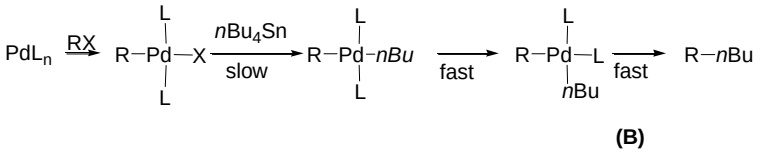
The catalytic cycle in **scheme 3.21** serves as a working model for the direct coupling reaction, although this cycle has yet to be firmly established, many of the individual steps in the cycle have been documented.[50]



Scheme 3.21: Mechanism of the catalyzed, direct coupling.[50]

Notably absent from the list of electrophiles in **table 3.2** are those organic compounds with a hydrogen on an sp^3 carbon in the β-position to the carbon bearing the leaving group. This limitation is imposed because the slowest step in the catalytic cycle, the transmetalation step, is much slower than the β-elimination that takes place in the alkylpalladium halide complex (**scheme 3.22 (A)**). This limitation is not imposed on the organotin partner; however, since the catalytic steps taking place after transmetalation – *trans*/*cis* isomerization and reductive elimination – are faster than β-elimination (**scheme 3.22 (B)**). Nevertheless, the process of the direct palladium-catalyzed coupling reactions of acid chlorides, benzyl halides, allyl halides, vinyl halides, vinyl triflates, and aryl halides with various organotin reagents have been reasonably well worked out.



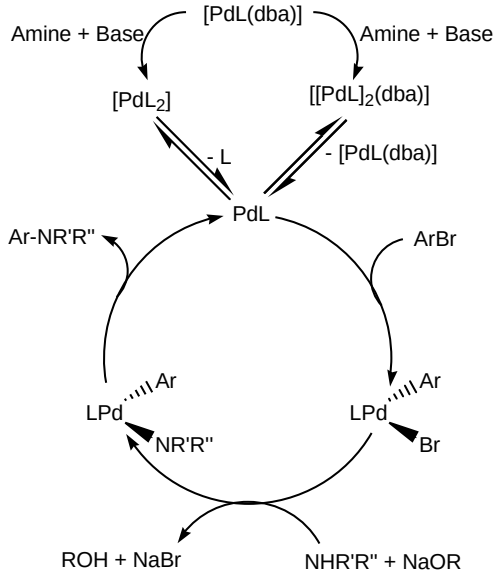


Scheme 3.22: (A) β-elimination in the alkylpalladium halide complex. (B) *trans*/*cis* isomerization and reductive elimination.[50]

3.6.9. Buchwald-Hartwig cross-coupling of acyclic amines with aryl triflates

The synthesis of compounds containing the *N*-aryl moiety has attracted a great deal of interest due to the importance of such compounds in different fields. Although a number traditional methods exist for aryl C-N bond construction, they typically require harsh conditions, the need to employ stoichiometric quantities of valuable reagents, numerous steps, or regiochemical ambiguities. Buchwald's[84-87] and Hartwig's groups[88] have progressively and independently contributed to the development of a general, reliable, and practical methodology for the formation of aromatic C-N bonds.

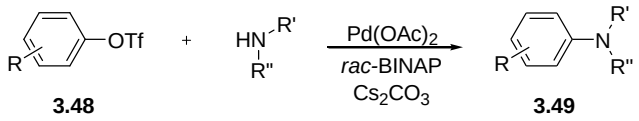
The catalytic cycle in **scheme 3.23** serves as a working model for the direct coupling reaction. This cycle has been firmly established after some controversy between both research groups but the agreement came in a joint publication in 2006 [89]. All of the individual steps in the cycle have been documented and established.[89,90]



Scheme 3.23: Buchwald-Hartwig cross-coupling.[89,90]

Using conditions similar to those developed for the amination of aryl bromides, both these groups found that good yields could be obtained utilizing

electronically neutral or electron-rich aryl triflates.[86] The amination of aryl-triflates was obtained using Cs_2CO_3 as a base in lieu of NaO*t*-Bu (**scheme 3.24**).[88]



3.48 **3.49**

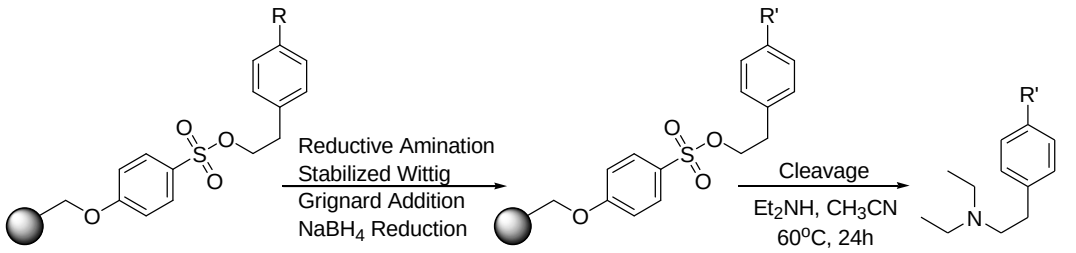
Scheme 3.24: Buchwald-Hartwig cross-coupling of acyclic amines with aryl triflates.[88]

3.6.10. Designing a "Traceless" tetrafluoroarysulfonate linker

Most linkers leave a residue attached to the cleaved molecule; that is a resin the functional group (or a derivative thereof) used to attach the linker to the resin is released as a different functional group (i.e. a carboxylic is released as a carboxamide, for example). Since this is not always desirable, there was interest in investigating linkers that left no obvious residue on the cleaved molecule. So-called "traceless linkers" are defined as ones where a new carbon-hydrogen, carbon-carbon or carbon-nitrogen bond is formed at the linkage site of the cleaved molecule. These linkers show non-specific function after cleavage and are so-called because an examination of the final compound reveals no trace of the point of linkage to the solid phase. In this specific case I was interested in perfluoroarylsulfonic linkers with palladium catalyzed reactions that could form these kinds of bonds as well as observing the behaviour of these reactions under microwave irradiation.

3.6.10.1. "Traceless" synthesis with supported aryl sulfonates

Reitz *et al*.[91,92] in 1998 reported the use of arylsulfonate esters in solid phase organic synthesis (SPOS). They showed that after linkage of alcohols to the resin, diversification by functional group interconversion was possible (**scheme 3.25**).

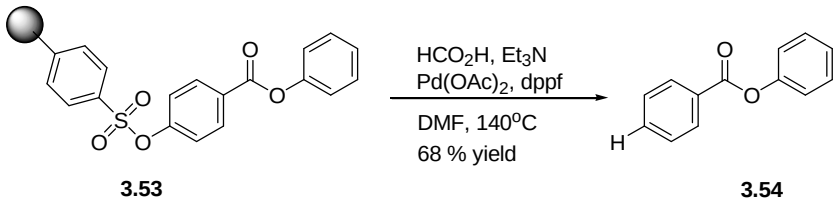


Resin 3.50	Reaction conditions	R' 3.52 a-m	Yield (%)	NMR Purity (%)
a	PhMgBr (5 eq), THF, 0°C, 4.5 h	-CH(OH)Ph	50	90
a	MeMgBr (5 eq), THF, 0°C, 4.5 h	-CH(OH)Me	42	90
a	$NaBH_4$ (10 eq), THF:MeOH (1:1), 30 m	$-CH_2OH$	66	90
a	$Ph_3P=CHCO_2Me$, (5 eq), THF, 70°C, 3h	-*E*-$CH=CHCO_2Me$	60	80
a	(4-MeO)$PhCH_2NH_2$ (7 eq), $NaBH(OAc)_3$ (7 eq), HOAc, CH_2Cl_2, 24 h	$-CH_2NHCH_2$[(4-MeO)Ph]	>95	85
a	(4-MeO)$PhNH_2$ (7 eq), $NaBH(OAc)_3$ (7 eq), HOAc, CH_2Cl_2, 24 h	$-CH_2NH$[(4-MeO)Ph]	53	95
a	$PhNH_2$ (7 eq), $NaBH(OAc)_3$ (7 eq), HOAc, CH_2Cl_2, 24 h	$-CH_2NHPh$	53	95
a	(4-F)$PhNH_2$ (7 eq), $NaBH(OAc)_3$ (7 eq), HOAc, CH_2Cl_2, 24 h	$-CH_2NH$[(4-F)Ph]	54	95
b	4-(MeO)PhCOCl (5 eq), $^iPr_2EtN$ (5 eq), CH_2Cl_2, 24 h	-NHC(O)[(4-MeO)Ph]	70	90
b	4-(NO_2)PhCOCl (5 eq), $^iPr_2EtN$ (5 eq), CH_2Cl_2, 24 h	-NHC(O)[(4-NO_2)Ph]	69	70
b	4-(MeO)$PhSO_2Cl$ (2 eq), $^iPr_2EtN$ (2 eq), CH_2Cl_2, 24 h	$-NHSO_2$[(4-MeO)Ph]	53	85
b	$PhCH_2NCO$ (5 eq), CH_2Cl_2, 24 h	$-NHC(O)NHCH_2Ph$	61	95
b	4-(MeO)PhNCO (5 eq), CH_2Cl_2, 24 h	-NHC(O)NH[(4-MeO)Ph]	45	90

Scheme 3.25: Reaction compatibility of resin –bound arylsulfonate esters followed by cleavage with Et_2NH (reproduced with permission).[91,92]

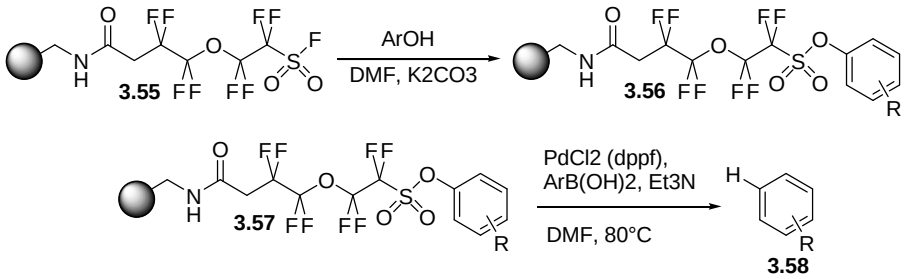
Unlike other methods in which alcohols are immobilized onto a polymer support, Reitz *et al.*[91,92] also observed that the cleavage step could be used to transform the original hydroxyl group into various useful functionalities offering the advantage of adding an element of diversity. At the same time Wustrow *et al.*[93] demonstrated the use of palladium (0) mediated reductive cleavage from

immobilized aryl sulfonates. Sulfonyl chloride functionalized polystyrene derivatised with 4-hydroxybenzoate was further elaborated at the carboxylate function, providing a small series of sulfonate ester amide linked compounds. Traceless liberation of the products was brought about by palladium catalyzed transfer hydrogenation effectively replacing the arylsulfonate linkage with a new aryl C-H bond. This methodology, however, was observed to suffer several drawbacks such as the need for high temperatures, long reaction times and the need for electron deficient substrates for efficient cleavage to occur, placing severe limitations on the range of chemistries possible (**scheme 3.26**).



Scheme 3.26: Reductive cleavage using the Wustrow linker.[93]

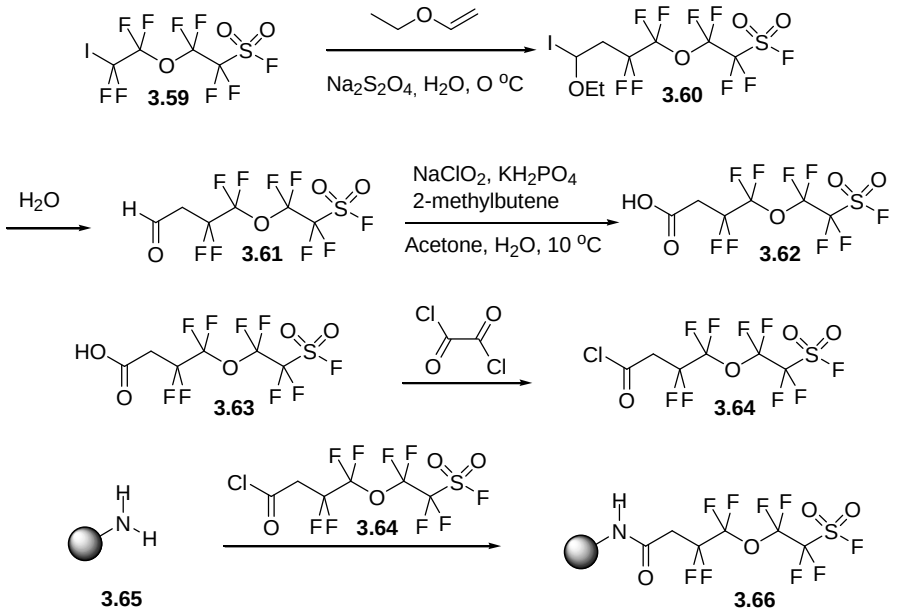
Holmes *et al.*[94] reported the synthesis of a traceless perfluoroalkylsulphonyl (PFS) linker for the deoxygenation of phenols **3.55**. Phenols were reacted with the perfluorosulphonyl fluoride group of the linker under mild conditions to give the stable arylsulphonate **3.56** tethered to the resin matrix via the acyl chloride chain (**scheme 3.28**). The polymer bound arylsulphonates were efficiently cleaved under mild reducing conditions to afford good yields of the deoxygenated aryl compounds **3.58**. Holmes also reported that this method for phenolic deoxygenation was very tolerant of a broad range of different functional groups, and therefore of considerable versatility for the synthesis of combinatorial libraries (**scheme 3.27**).



3.55 **3.56** **3.57** **3.58**

Scheme 3.27: Derivatization of phenolic substrates and transfer hydrogenation using the Holmes linker.[94]

The PFS linker designed by Holmes *et al.*[94] was synthesized through a five step route, commencing from an expensive, commercially available, iodide **3.59** (**scheme 3.28**).



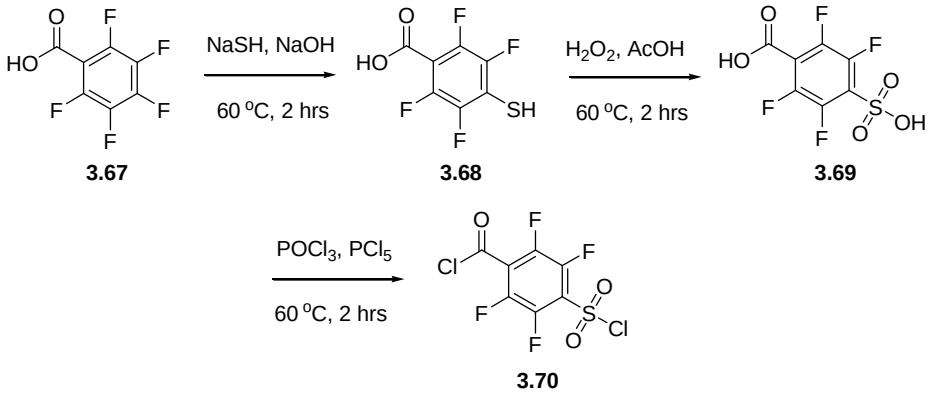
3.59 **3.60** **3.61** **3.62** **3.63** **3.64** **3.65** **3.66**

Scheme 3.28: Synthesis of the Holmes perfluoroalkylsulfonate linker.[94]

3.6.10.2. Preparation of an immobilized tetrafluoroarylsulfonyl chloride "traceless" linker

Inspired by the results obtained by Holmes, and considering the linkers used by Reitz and Wustrow, the Combinatorial Centre of Excellence[21] (CCE) proposed a conceptually similar linker albeit based around a tetrafluorophenylsulphonyl moiety.

This linker synthesis commenced from commercially available pentafluorobenzoic acid **3.72**, and followed a literature procedure[95] for the preparation of the mixed sulphonyl/carbonyl chloride **3.75** (**scheme 3.29**).



3.67 **3.68** **3.69** **3.70**

Scheme 3.29: Synthesis of 4-chlorosulfonyl-2,3,5,6-tetraflourobenzoyl chloride.[96,97]

Pentafluorobenzoic acid **3.67** reacted under aqueous basic conditions with the highly nucleophilic sodium hydrosulphide in a S_NAr reaction to give the tetrafluorobenzothiol **3.68**. Oxidation to the sulphonic acid using hydrogen peroxide in acetic acid afforded the sulphonic acid derivative **3.69**. While chlorination of both the carboxylic and sulphonic acid was performed using a mixture of phosphorus pentachloride and phosphorus trichloride. Purification of the mixed acid chloride **3.70** however proved difficult, and removal of traces of orthophosphoric acid and phosphorus oxychloride proved impossible unless high vacuum distillation was carried out, since at temperatures exceeding 150°C significant decomposition was observed.[96,97]

With these advances in mind the next step to take is to prepare the "traceless" perfluoroarylsulfonate linker **3.70** and test it under microwave conditions. In the next chapter the approaches using microwave heating conditions and the results are presented and discussed.

3.7. References

(1) Hunter, D. *Journal of Cellular Biochemistry* **2001**, 22-27.

(2) Tzschucke, C. C.; Markert, C.; Bannwarth, W.; Roller, S.; Hebel, A.; Haag, R. *Angewandte Chemie-International Edition* **2002**, *41*, 3964-4000.

(3) Kappe, C. O. *Angewandte Chemie-International Edition* **2004**, *43*, 6250-6284.

(4) Lidstrom, P.; Tierney, J.; Wathey, B.; Westman, J. *Tetrahedron* **2001**, *57*, 9225-9283.

(5) Olofsson, K.; Hallberg, A.; Larhed, M. In *Microwaves in Organic Synthesis*; Loupy, A., Ed.; Wiley-VCH: Weinheim, 2002, p 379.

(6) Gladysz, J. A.; Curran, D. P. *Tetrahedron* **2002**, *58*, 3823-3825.

(7) Curran, D. P. In *Handbook of Fluorous Chemistry*; Gladysz, J. A., Curran, D. P., Horvath, I. T., Eds.; Wiley-VCH: Weinheim, 2004, p 128.

(8) Studer, A.; Hadida, S.; Ferritto, R.; Kim, S. Y.; Jeger, P.; Wipf, P.; Curran, D. P. *Science* **1997**, *275*, 823-826.

(9) Chen, C. H. T.; Zhang, W. *Organic Letters* **2003**, *5*, 1015-1017.

(10) Bose, A. K.; Manhas, M. S.; Ganguly, S. N.; Sharma, A. H.; Banik, B. K. *Synthesis-Stuttgart* **2002**, 1578-1591.

(11) Caddick, S. *Tetrahedron* **1995**, *51*, 10403-10432.

(12) Elander, N.; Jones, J. R.; Lu, S. Y.; Stone-Elander, S. *Chemical Society Reviews* **2000**, *29*, 239-249.

(13) Loupy, A. *Microwaves in Organic Synthesis*; Wiley-VCH: Weinheim, 2002.

(14) Mingos, D. M. P.; Baghurst, D. R. *Chemical Society Reviews* **1991**, *20*, 1-47.

(15) Strauss, C. R.; Trainor, R. W. *Australian Journal of Chemistry* **1995**, *48*, 1665-1692.

(16) Larhed, M.; Hallberg, A. *Drug Discovery Today* **2001**, *6*, 406-416.

(17) Kaiser, N. F. K.; Hallberg, A.; Larhed, M. *Journal of Combinatorial Chemistry* **2002**, *4*, 109-111.

(18) Alterman, M.; Andersson, H. O.; Garg, N.; Ahlsen, G.; Lovgren, S.; Classon, B.; Danielson, U. H.; Kvarnstrom, I.; Vrang, L.; Unge, T.; Samuelsson, B.; Hallberg, A. *Journal of Medicinal Chemistry* **1999**, *42*, 3835-3844.

(19) Ersmark, K.; Feierberg, I.; Bjelic, S.; Hamelink, E.; Hackett, F.; Blackman, M. J.; Hulten, J.; Samuelsson, B.; Aqvist, J.; Hallberg, A. *Journal of Medicinal Chemistry* **2004**, *47*, 110-122.

(20) Kappe, C. O. *Current Opinion in Chemical Biology* **2002**, *6*, 314-320.

(21) Bradley, M. *Current Medicinal Chemistry* **2002**, *9*, 2173-2177.

(22) Lew, A.; Krutzik, P. O.; Hart, M. E.; Chamberlin, A. R. *Journal of Combinatorial Chemistry* **2002**, *4*, 95-105.

(23) Gabriel, C.; Gabriel, S.; Grant, E. H.; Halstead, B. S. J.; Mingos, D. M. P. *Chemical Society Reviews* **1998**, *27*, 213-223.

(24) Stadler, A.; Kappe, C. O. *Journal of Combinatorial Chemistry* **2001**, *3*, 624-630.

(25) Noteberg, D.; Hamelink, E.; Hulten, J.; Wahlgren, M.; Vrang, L.; Samuelsson, B.; Hallberg, A. *Journal of Medicinal Chemistry* **2003**, *46*, 734-746.

(26) Schaal, W.; Karlsson, A.; Ahlsen, G.; Lindberg, J.; Andersson, H. O.; Danielson, U. H.; Classon, B.; Unge, T.; Samuelsson, B.; Hulten, J.; Hallberg, A.; Karlen, A. *Journal of Medicinal Chemistry* **2002**, *45*, 752-752.

(27) Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T.; Hulten, J.; Bonham, N. M.; Schaal, W.; Karlen, A.; Hallberg, A. *Journal of Medicinal Chemistry* **1997**, *40*, 898-902.

(28) Ax, A.; Schaal, W.; Vrang, L.; Samuelsson, B.; Hallberg, A.; Karlen, A. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 755-764.

(29) Alterman, M.; Bjorsne, M.; Muhlman, A.; Classon, B.; Kvarnstrom, I.; Danielson, H.; Markgren, P. O.; Nillroth, U.; Unge, T.; Hallberg, A.; Samuelsson, B. *Journal of Medicinal Chemistry* **1998**, *41*, 3782-3792.

(30) Wannberg, J.; Kaiser, N. F. K.; Vrang, L.; Samuelsson, B.; Larhed, M.; Hallberg, A. *Journal of Combinatorial Chemistry* **2005**, *7*, 611-617.

(31) Wannberg, J.; Larhed, M. *Journal of Organic Chemistry* **2003**, *68*, 5750-5753.

(32) Crosignani, S.; White, P. D.; Linclau, B. *Journal of Organic Chemistry* **2004**, *69*, 5897-5905.

(33) McAllister, L. A.; McCormick, R. A.; Brand, S.; Procter, D. J. *Angewandte Chemie-International Edition* **2005**, *44*, 452-455.

(34) El Bakkari, M.; Vincent, J. M. *Organic Letters* **2004**, *6*, 2765-2767.

(35) Curran, D. P.; Fischer, K.; Moura-Letts, G. *Synlett* **2004**, 1379-1382.

(36) Herrero, M. A.; Wannberg, J.; Larhed, M. *Synlett* **2004**, 2335-2338.

(37) Cioffi, C. L.; Berlin, M. L.; Herr, R. J. *Synlett* **2004**, 841-845.

(38) Nagashima, T.; Zhang, W. *Journal of Combinatorial Chemistry* **2004**, *6*, 942-949.

(39) Ritter, K. *Synthesis-Stuttgart* **1993**, 735-762.

(40) Stang, P. J.; Hanack, M.; Subramanian, L. R. *Synthesis-Stuttgart* **1982**, 85-126.

(41) Voss, P.; Niederprum, H.; Beyl, V. *Liebigs Ann. Chem.* **1973**, *20*, 731-743.

(42) Yoneda, N.; Fukuhara, T.; Mizokami, T.; Suzuki, A. *Chemistry Letters* **1991**, 459-460.

(43) Dupre, B.; Meyers, A. I. *Journal of Organic Chemistry* **1991**, *56*, 3197-3198.

(44) Cacchi, S.; Morera, E.; Ortar, G. *Organic Syntheses* **1990**, *68*, 138-147.

(45) Engler, T. A.; Reddy, J. P.; Combrink, K. D.; Vandervelde, D. *Journal of Organic Chemistry* **1990**, *55*, 1248-1254.

(46) Saa, J. M.; Dopico, M.; Martorell, G.; Garciaraso, A. *Journal of Organic Chemistry* **1990**, *55*, 991-995.

(47) Lipshutz, B. H.; Buzard, D. J.; Vivian, R. W. *Tetrahedron Letters* **1999**, *40*, 6871-6874.

(48) Cacchi, S.; Ciattini, P. G.; Morera, E.; Ortar, G. *Tetrahedron Letters* **1986**, *27*, 5541-5544.

(49) Percec, V.; Bae, J. Y.; Hill, D. H. *Journal of Organic Chemistry* **1995**, *60*, 1060-1065.

(50) Stille, J. K. *Angewandte Chemie-International Edition* **1986**, *25*, 508-523.

(51) Scott, W. J.; Stille, J. K. *Journal of the American Chemical Society* **1986**, *108*, 3033-3040.

(52) Percec, V.; Bae, J. Y.; Zhao, M. Y.; Hill, D. H. *Journal of Organic Chemistry* **1995**, *60*, 176-185.

(53) Miyaura, N.; Suzuki, A. *Chemical Reviews* **1995**, *95*, 2457-2483.

(54) Miyaura, N.; Yanagi, T.; Suzuki, A. *Synthetic Communications* **1981**, *11*, 513-519.

(55) Ishiyama, T.; Kizaki, H.; Hayashi, T.; Suzuki, A.; Miyaura, N. *Journal of Organic Chemistry* **1998**, *63*, 4726-4731.

(56) Ishiyama, T.; Kizaki, H.; Miyaura, N.; Suzuki, A. *Tetrahedron Letters* **1993**, *34*, 7595-7598.

(57) Labadie, S. S. *Synthetic Communications* **1994**, *24*, 709-719.

(58) Hartwig, J. F. *Angewandte Chemie-International Edition* **1998**, *37*, 2047-2067.

(59) Cortese, N. A.; Ziegler, C. B.; Hrnjez, B. J.; Heck, R. F. *Journal of Organic Chemistry* **1978**, *43*, 2952-2958.

(60) Shen, W. *Tetrahedron Letters* **1997**, *38*, 5575-5578.

(61) Ashimori, A.; Matsuura, T.; Overman, L. E.; Poon, D. J. *Journal of Organic Chemistry* **1993**, *58*, 6949-6951.

(62) Muller, W.; Lowe, D. A.; Neijt, H.; Urwyler, S.; Herrling, P. L.; Blaser, D.; Seebach, D. *Helvetica Chimica Acta* **1992**, *75*, 855-864.

(63) Thompson, W. J.; Jones, J. H.; Lyle, P. A.; Thies, J. E. *Journal of Organic Chemistry* **1988**, *53*, 2052-2055.

(64) Shieh, W. C.; Carlson, J. A. *Journal of Organic Chemistry* **1992**, *57*, 379-381.

(65) Littke, A. F.; Fu, G. C. *Angewandte Chemie-International Edition* **1998**, *37*, 3387-3388.

(66) Watanabe, T.; Miyaura, N.; Suzuki, A. *Synlett* **1992**, 207-210.

(67) Gronowitz, S.; Bobosik, V.; Lawitz, K. *Chemica Scripta* **1984**, *23*, 120-122.

(68) Gronowitz, S.; Lawitz, K. *Chemica Scripta* **1984**, *24*, 5-6.

(69) Barn, D.; Caulfield, W.; Cowley, P.; Dickins, R.; Bakker, W. I.; McGuire, R.; Morphy, J. R.; Rankovic, Z.; Thorn, M. *Journal of Combinatorial Chemistry* **2001**, *3*, 534-541.

(70) Hanazawa, T.; Wada, T.; Masuda, T.; Okamoto, S.; Sato, F. *Organic Letters* **2001**, *3*, 3975-3977.

(71) Lutz, C.; Bleicher, K. H. *Tetrahedron Letters* **2002**, *43*, 2211-2214.

(72) Park, C.; Burgess, K. *Journal of Combinatorial Chemistry* **2001**, *3*, 257-266.

(73) Pourbaix, C.; Carreaux, F.; Carboni, B. *Organic Letters* **2001**, *3*, 803-805.

(74) Zhang, H. C.; Ye, H.; White, K. B.; Maryanoff, B. E. *Tetrahedron Letters* **2001**, *42*, 4751-4754.

(75) Fraley, M. E.; Rubino, R. S. *Tetrahedron Letters* **1997**, *38*, 3365-3368.

(76) Piettre, S. R.; Baltzer, S. *Tetrahedron Letters* **1997**, *38*, 1197-1200.

(77) Heck, R. F. *Organic Reactions* **1982**, *27*, 345-390.

(78) Chen, Q. Y.; Yang, Z. Y. *Tetrahedron Letters* **1986**, *27*, 1171-1174.

(79) Olofsson, K.; Larhed, M.; Hallberg, A. *Journal of Organic Chemistry* **1998**, *63*, 5076-5079.

(80) Demeijere, A.; Meyer, F. E. *Angewandte Chemie-International Edition* **1995**, *33*, 2379-2411.

(81) Lorsbach, B. A.; Kurth, M. J. *Chemical Reviews* **1999**, *99*, 1549-1581.

(82) Sammelson, R. E.; Kurth, M. J. *Chemical Reviews* **2001**, *101*, 137-202.

(83) Akaji, K.; Teruya, K.; Akaji, M.; Aimoto, S. *Tetrahedron* **2001**, *57*, 2293-2303.

(84) Marcoux, J. F.; Wagaw, S.; Buchwald, S. L. *Journal of Organic Chemistry* **1997**, *62*, 1568-1569.

(85) Wolfe, J. P.; Buchwald, S. L. *Journal of Organic Chemistry* **1997**, *62*, 1264-1267.

(86) Ahman, J.; Buchwald, S. L. *Tetrahedron Letters* **1997**, *38*, 6363-6366.

(87) Wolfe, J. P.; Tomori, H.; Sadighi, J. P.; Yin, J. J.; Buchwald, S. L. *Journal of Organic Chemistry* **2000**, *65*, 1158-1174.

(88) Louie, J.; Driver, M. S.; Hamann, B. C.; Hartwig, J. F. *Journal of Organic Chemistry* **1997**, *62*, 1268-1273.

(89) Shekhar, S.; Ryberg, P.; Hartwig, J. F.; Mathew, J. S.; Blackmond, D. G.; Strieter, E. R.; Buchwald, S. L. *Journal of the American Chemical Society* **2006**, *128*, 3584-3591.

(90) Biscoe, M. R.; Barder, T. E.; Buchwald, S. L. *Angewandte Chemie-International Edition* **2007**, *46*, 7232-7235.

(91) Rueter, J. K.; Nortey, S. O.; Baxter, E. W.; Leo, G. C.; Reitz, A. B. *Tetrahedron Letters* **1998**, *39*, 975-978.

(92) Baxter, E. W.; Rueter, J. K.; Nortey, S. O.; Reitz, A. B. *Tetrahedron Letters* **1998**, *39*, 979-982.

(93) Jin, S. J.; Holub, D. P.; Wustrow, D. J. *Tetrahedron Letters* **1998**, *39*, 3651-3654.

(94) Pan, Y. J.; Holmes, C. P. *Organic Letters* **2001**, *3*, 2769-2771.

(95) Fielding, H. C.; Shirley, I. M. *Journal of Fluorine Chemistry* **1992**, *59*, 15-31.

(96) Revell, J. D. PhD, Southampton, 2003.

(97) Revell, J. D.; Ganesan, A. *Chemical Communications* **2004**, 1916-1917.

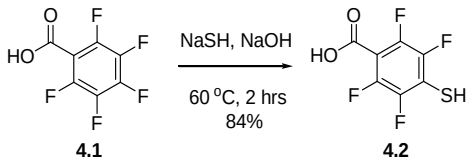
Chapter Four

Synthesis and optimization of a linker for Microwave-assisted Palladium chemistry

The purpose of this chapter is to report the different approach on the synthesis of a perfluoroarylsulfonate linker designed for palladium mediated chemistries as an alternative to the Holmes linker. The different optimization procedures for the different palladium-mediated chemistries and results obtained are described and discussed.

4. Preparation of the "traceless" linker

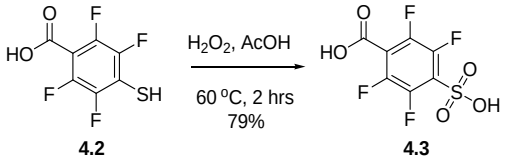
The preparation of 4-chlorosulfonyl-2, 3, 5, 6-tetrafluorobenzoyl chloride was published by Fielding *et al.*[1] in 1992, and was considered to provide an ideal linker platform for both reductive cleavage and cross-coupling chemistries. Synthesis of the linker started with the S_NAr reaction between pentafluorobenzoic acid **4.1** and sodium hydrosulfide (NaSH) under strongly basic aqueous conditions (**scheme 4.1**). As noted by Fielding *et al.*[1], the stoichiometry and rate of addition of pentafluorobenzoic acid **4.1** to the basic mixture had to be closely controlled to avoid formation of a thioether side-product. Nevertheless, the side product could be removed, since it presents itself as a supernatant green particle suspension after neutralization with HCl and this supernatant layer can be discarded.



4.1 **4.2**

Scheme 4.1 – Preparation of 4-mercapto-2,3,5,6-tetrafluorobenzoic acid 4.2.[1]

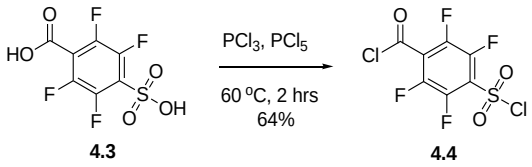
Oxidation of 4- mercaptotetrafluorobenzoic acid **4.2** to the bifunctional mixed acid **4.3**, was accomplished using hydrogen peroxide in acetic acid (**scheme 4.2)**. Close temperature regulation over the rate of addition of the substrate to the oxidizing mixture was required since the oxidation itself is highly exothermic. Following oxidation, excess of hydrogen peroxide was destroyed by addition of sodium metabisulfite, and the crude product is isolated by liquid-liquid extraction into ethyl acetate. Back extraction of the bis-acid **4.3** into water allowed isolation of the product after azeotropic removal of water. Removal of inorganic salts was difficult given the high solubility of the product in both water and organic solvents. Since the removal of salts was a difficult process which led to a significance loss of product, two different workups were tried to minimise such loss. In the first, the removal of salts was made using an ion exchange resin (Amberlite 200). In the second, extending the reflux for up to 4 hours allowed the decomposition of the peroxide to occur without the need to add sodium metabisulfite. Nevertheless, in this second process it must be taken into account that no decomposition of the peroxide is made of the peroxide present are difficult to identify by mass spectrometry or HPLC. These traces require that in the evaporation process, the evaporator system and pump must not have any metal parts, and therefore a membrane pumps rather than oil pump was used.



Scheme 4.2 – Preparation of 4-sulpho-2,3,5,6-tetrafluorobenzoic acid 4.3.[1]

Thorough drying of the mixed acid is essential prior to halogenation by reaction with PCl_3 and PCl_5 (**scheme 4.3**). This was achieved by preliminary drying at 90°C in a vacuum oven over 24 hours, followed by removal of traces of residual water by desiccation over P_2O_5 *in vacuo*, for a further 24 hours. Halogenation was achieved by addition of the anhydrous bis-acid **4.3** to a mixture of PCl_5 and PCl_3 (10 eq. in 1:4 w/v ratio) at 60°C with rapid stirring under argon for 4 hours. Hydrogen

chloride gas produced was vented by way of an oil bubbler (and subsequently quenched) to maintain an anhydrous atmosphere over the reaction mixture. Removal of excess phosphorus reagents and by-products by low-pressure distillation provided the crude bis-acid chloride **4.4**. High vacuum Kügelrhor distillation gave the pure desired compound as fuming moisture sensitive colourless liquid. Occasionally the bis-acid chloride **4.4** distilled as a yellow liquid, which despite colouration, reacted in an identical manner to the colourless substance.

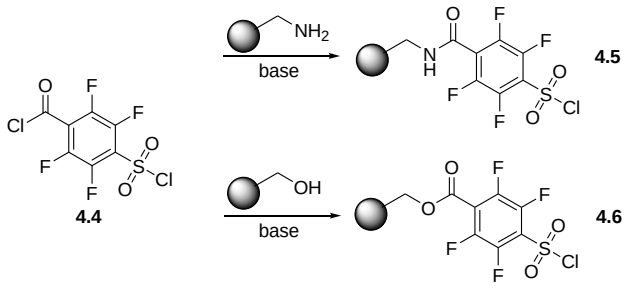


4.3 **4.4**

Scheme 4.3 - Preparation of 4-chlorosulphonyl-2,3,5,6-tetrafuorobenzoyl chloride 4.4.

4.1. Preparation of an immobilized tetrafluoroarylsulfonyl chloride.

The significant difference in reactivity between the sulfonyl and acyl chloride groups of **4.4** identified by Fielding[1] was exploited to advantage, allowing selective functionalization of the acyl chloride at the first stage (**scheme 4.4**).



4.4 **4.5** **4.6**

Scheme 4.4: Reaction of bis-acid chloride 4.4 with aminomethyl- and hydroxymethyl polystyrene.

Both the polystyrene immobilised amide and ester linked sulfonyl chlorides **4.5** and **4.6** have been prepared previously, and are termed 2,3,5,6-

tetrafluorobenzamido-4-sulfonyl chloride-polystyrene and 2,3,5,6-tetrafluorobenzoyloxymethyl-4-sulfonyl chloride-polystyrene resin respectively. The inventions are listed within the patent **WO 99/67228** published by Salvino, Joseph, M (Inventor), (29.12.99), on behalf of Rhone-Poulenc. These fluorophenyl resin inventions, their respective methods of preparation, and also their use in the solid phase synthesis of amides, peptides, hydroxamic acids, amines, urethanes, carbonates, carbamates, sulfonamides and α-substituted carbonyl compounds are protected.

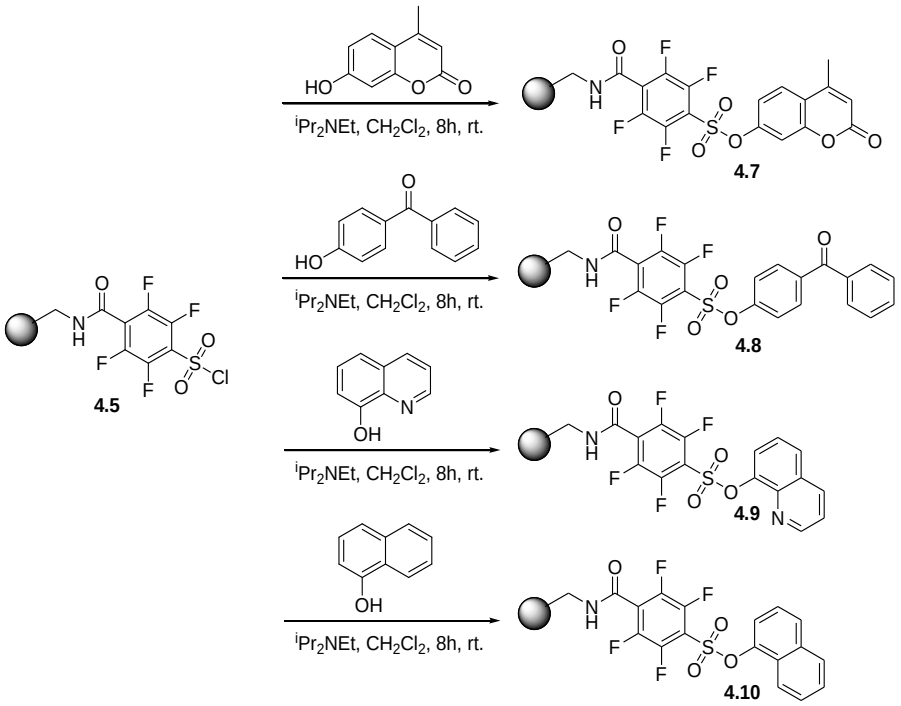
Our preparation and intended use of these resins is however significantly different from that disclosed within the aforementioned patent. Reaction of the bis-acid chloride **4.4** with aminomethyl functionalised polystyrene in the presence of a non-nucleophilic base allowed derivatisation of the support via the amide or ester respectively (**scheme 4.4**). The hydroxymethyl functionalised polystyrene version of the linker was not used since studies carried out by Revell[2] indicated better results when using the aminomethyl functionalised version.

Addition of the bis-acid chloride **4.4** in dichloromethane invariably resulted in highly coloured resins even after extensive washing procedures when compared to the inverse addition. In the cases of tetrabutylammonium hydroxide and tetrabutylammonium hydrocarbonate, addition of the bis-acid chloride gave rise to an exothermic reaction and rapid colouration in both dichloromethane and dimethylformamide

Inverse addition of a dilute solution of the base to a mixture of the bis-acid chloride and resin slurred in dichloromethane was found to give far superior results. The most satisfactory results were obtained by preliminary addition of the bis-acid chloride to the resin in anhydrous dichloromethane at $0^{o}C$ followed by agitation at this temperature for 1 hour. To the cooled reaction mixture was added, drop wise over 30 minutes, a dilute solution of diisopropylethylamine in dichloromethane (10% v/v) maintaining the temperature and vigorous agitation throughout the addition.

4.2. Sulfonate formation by reaction with phenols and alcohols

With small batches of AMP **4.5** tetrafluorosulfonyl chloride resins in hand, formation of sulfonate derivatives was pursued by reaction with several phenolic compounds in the presence of a base (**scheme 4.5**). For initial studies, four readily available substrates were chesen for derivatisation: 7-hydroxy-4-methylcoumarin, 4-hydroxybenzophenone, 8-hydroxyquinoline and 8-hydroxynaphthalene.



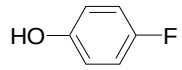
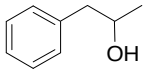
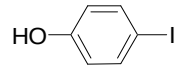
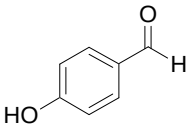
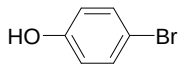
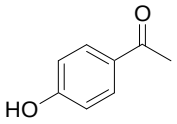
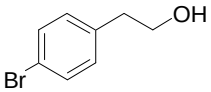
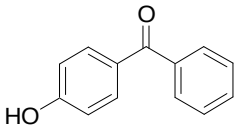
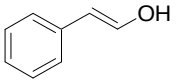
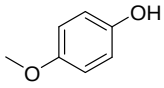
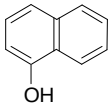
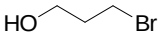
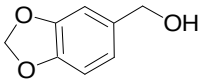
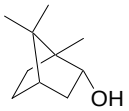
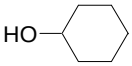
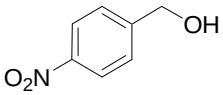
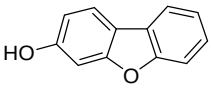
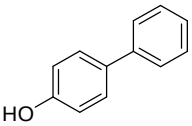
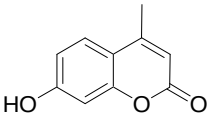
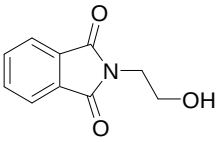
Scheme 4.5 - Phenolyc derivatisation using the tetrafluoroarylsulfonyl chloride linker.

Each substrate was dissolved in a small quantity of dichloromethane, and added in one portion to a slurry of the sulfonyl chloride resin **4.5**, pre-swelled in dichoromethane. In the case of 7-hydroxy-4-methylcoumarin, which is only sparingly soluble in dichloromethane, a minimum volume of dimethylformamide was added to aid dissolution. A 3-fold excess of Hünig's base was added, and the reaction mixture was agitated at room temperature for 8 hours, providing after thorough washing, the respective immobilised sulfonate ester.

The resin was reacted with different phenols and alcohols in order to verify its different reactivities and scope (as a possible scavenger) for these functional

groups, besides its use in cross-coupling reactions. The resin bound sulfonates loadings were determined (mainly) by two different methods: the first consisted in a direct weighting of the resin bound substrates and direct mass quantification, the second in cleaving the bound substrates using a fluoride source (see **section 4.3**), purification and weight determination of these sulfonates. Typical loading of 0.75mmolg^{-1} was determined for these substrates was obtained and verified by elemental analysis for one of the immobilised sulfonates prepared in three different assays (**table 4.1**).

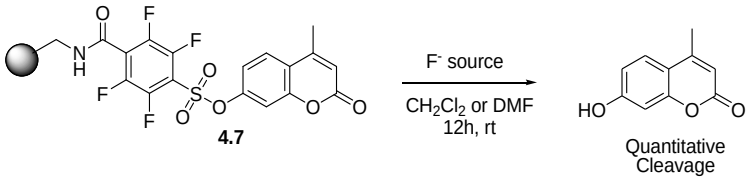
Table 4.1 - Different alcohols used for sulfonates formation and respective loadings.

Substrate	Loading (mmol/g)	Entry	Substrate	Loading (mmol/g)	Entry
	0.77	**4.11**		0.83	**4.21**
	0.52	**4.12**		0.85	**4.22**
	0.48 0.46*	**4.13**		0.71	**4.23**
	0.46	**4.14**		0.68	**4.24**
	0.65	**4.15**		0.64	**4.25**
	0.74	**4.16**		0.50	**4.26**
	0.84	**4.17**		0.72	**4.27**
	0.70	**4.18**		0.84	**4.28**
	0.72	**4.19**		0.72	**4.29**
	0.49	**4.20**		0.79	**4.30**

All loadings determined by weight of the bounded compound; (*) was determined by elemental analysis in three different assays (same value of loading was obtained in all assays).

4.3. Reaction of immobilised sulfonates with a fluoride source

As part of an investigation into selective fluorination of aromatic compounds, the reaction of immobilised tetrafluorosulfonate esters with a fluoride source was hypothesised as a possible route This assessment had been made taking into account the similarity of the tetrafluoroarylsulfonyl chloride linker with the Wustrow linker.[20] However, reaction of 7-hydroxy-4-coumarin immobilised as the tetrafluorosulfonate ester **4.7** with TBAF (aq), ammonium fluoride (anhydrous) or sodium fluoride invariably led to hydrolytic cleavage of the S-O bond of the sulfonate group, and quantitative isolation of 7-hydroxy-4-coumarin. Hence although this method is not suitable for introduction of a fluoride atom in place of the sulfonate ester, it does serve as a useful method for the quantitative analysis of substrate loading following functional groups manipulation.



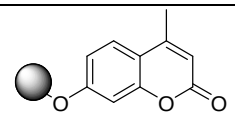
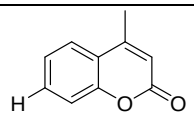
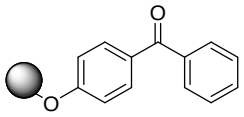
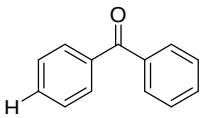
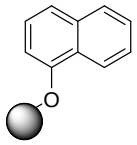
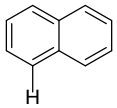
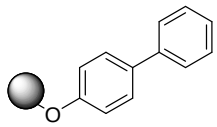
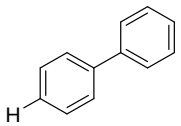
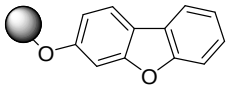
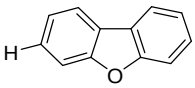
Scheme 4.5: Quantitative cleavage of 7-hydroxy-4-methylcoumarin by a fluoride source.

4.4. Transfer hydrogenation of immobilised sulfonates

In order to evaluate the reactivity of these immobilised sulfonate esters under conditions of transfer hydrogenation, the standardised procedure adopted by Holmes[3] had been employed by Revell[2,4] with exciting results. The supported sulfonate esters were individually pre-swollen in anhydrous DMF, which was degassed with nitrogen immediately prior to use. A mixture of 10% mol $Pd(OAc)_2$ and 15% dppp was found to act as an adequate Pd/ligand system, and in the presence of an excess of anhydrous formic acid and triethylamine at 80°C all immobilised sulfonate esters were reduced to the des-hydroxyaryl compounds in good to excellent yields by Revell (**table 4.2**).[2] In order to reduce the reaction time, a microwave assisted procedure for the transfer

hydrogenation was carried out at 180°C for a period of 10 minutes (the increase in pressure at this combination was found to be less dangerous than the first attempted at 220°C for 5 minutes). In all cases, the reduced aromatic product were obtained in the same range of yields as reported by Revell (**table 4.2**).[2,4]

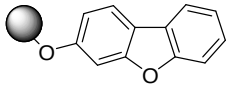
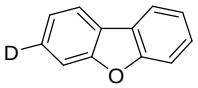
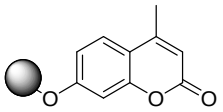
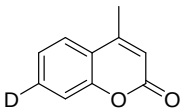
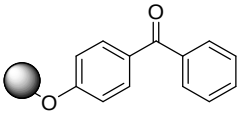
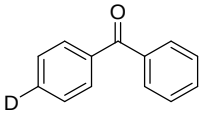
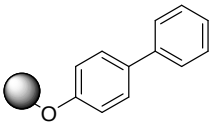
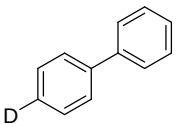
Table 4.2 – Isolated yields for transfer hydrogenation of sulfonate esters using microwave and reported for normal heating[2,4].

Solid supported compound	Reductive cleavage product	Entry	Yield (%)	Yield (%)[2,4]
		**4.31**	76	82
		**4.32**	84	89
		**4.33**	75	80
		**4.34**	72	79
		**4.35**	70	81

4.5. Transfer deuteration of immobilised sulfonates

Given the excellent yields obtained using the tetrafuoroarylsulfonate linker under transfer hydrogenation conditions, similar yields of deuterated aromatic compounds were anticipated on exchanging HCO_2H for DCO_2D. In order to avoid any possibility of proton exchange as reported by Revell[2,4], DMA was used instead of DMF as the reation solvent. Under otherwise identical reaction conditions ($Pd(OAc)_2$/dppf and triethylamine) substrates were hydrogenated with excellent isolated yields of the isotopically labelled deuterioaryl compounds (transfer deuteration was carried out at 180°C for a period of 15 minutes, **table 4.3**).

Table 4.3 -Transfer deuteration cleavage products.

Solid supported compound	Reductive deuteration product	Entry	Yield (%)	Yield (%)[2,4]
		**4.36**	78	76
		**4.37**	80	80
		**4.38**	81	89
		**4.39**	72	77

4.6. Optimisation of conditions for Suzuki-Miyaura cross-coupling

In order to optimise the isolated yields obtained using the AMP immobilised sulfonates, several different catalyst/ligand systems were evaluated in the cross-coupling of phenylboronic acid with the derivatised 7-hydroxy-4-methylcoumarin. Reactions were carried out under microwave irradiation at 180^{o}C for a period of 30 minutes and samples were analysed. However, in keeping with findings reported by Holmes[3] and Revell[2], $Pd(OAc)_2$/dppp, $Pd(OAc)_2/PPh_3$, $PdCl_2(PPh_3)_2$ and $Pd_2(dba)_3$ all resulted in lower yields in conjunction with triethylamine and phenylboronic acid, and $PdCl_2$(dppf) was the best catalyst (**table 4.4**).

Table 4.4: Isolated yields of 7-phenylmethylcoumarin using various different catalyst/ligand systems under microwave irradiation.

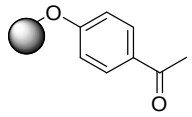
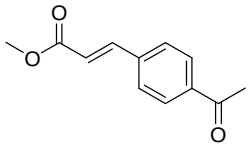
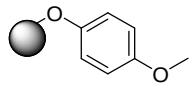
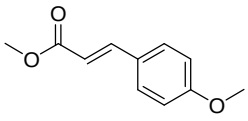
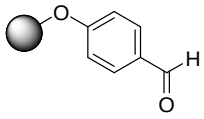
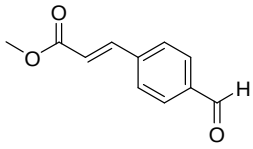
Entry number	Catalyst/ligand system	Isolared Yield (%)
1	$Pd(OAc)_2$/dppp	60
2	$Pd(OAc)_2/PPh_3$	48
3	$PdCl_2(PPh_3)_2$	66
4	$Pd_2(dba)_3$	63
5	$PdCl_2$(dppf)	78

Since all the findings had similar results when compared to the reactions performed by Revell[2] no further optimizations were carried out.

4.7. Heck cross-coupling with methyl acrylate

Possibilities available for use with this linker also extend to the reaction of immobilised sulfonates under Heck cross-coupling, in the presence of a suitable catalyst and methyl acrylate. Experiments had been conducted using 10 mol % $PdCl_2$(dppf) with 10 equivalents of triethylamine in dry DMF by Revell.[2] Heating of immobilised (4-hydroxybenzaldehyde)sulfonate resin **4.22** to 100, 130 and 160^oC for 30 minutes in the presence of 10 equivalents of methyl acrylate, at 160^oC for a period of 35 minutes under nitrogen atmosphere and microwave irradiation, produced a very complicated crude reaction mixture by TLC analysis. For this reason, and following the conclusions of Revell[2], 2 molar equivalents of Bu_4NI were added and the reaction was repeated, and pleasingly the desired products were obtained after preparative TLC of the concentrated crude reaction mixture following aqueous workup and extraction of the products into diethyl ether. The results obtained are presented in **table 4.5**.

Table 4.5: Heck cross-coupling reaction products under microwave irradiation.

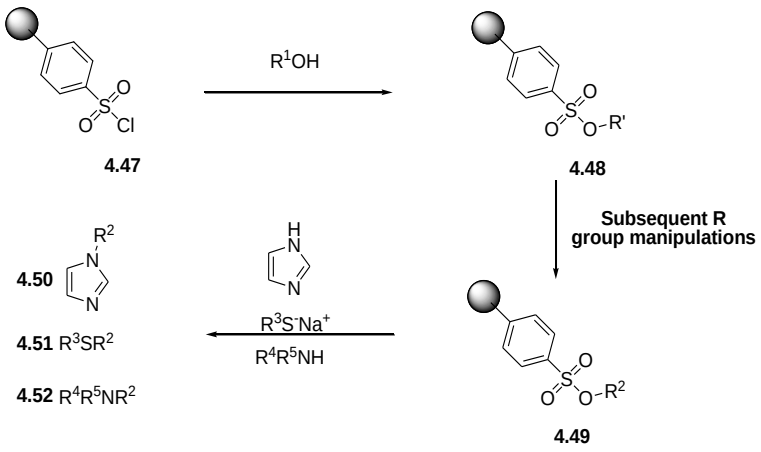
Solid supported compound	Heck reaction products*	Entry	Yield (%)
		**4.40**	76
		**4.41**	61
		**4.42**	63

*Spectroscopic data for products formed in these reactions were identical with those reported previously[2,4].

4.8. Results and Discussion

The synthesis and immobilisation of 4-chlorosulfonyl-2,3,5,6-tetrafluorobenzoylchloride **4.4** onto aminomethyl polystyrene resin through formation of the amide **4.5** as been demonstrated. The versatility of this "traceless" linker has been shown in the different results on the immobilization of phenolic or alcohol compounds as sulfonate esters, but most importantly for transfer hydrogenation and deuteration, as well as for Suzuki-Miyaura and Heck cross-coupling reactions under microwave irradiation. The optimization of the microwave conditions for these previously reported[2,4] reactions provides an important tool for the fast production of possible inhibitors or libraries of small compounds, since the reactivity of the linker is not limited to the mere scavenging or palladium reactive roles but also it is possible to cleave a desired substrate at any step by using a fluoride source.

In the cases of the Buchwald-Hartwig and Stille cross-couplings the possibility of further increasing the range of possibilities for the application of the linker was not successful. The possibility of using different conditions, using different palladium sources, different ligands or different combinations of solvents remains unscreened. One direction could be the use of ionic liquids as solvents, as their known to present good properties under microwave irradiation may provide different outcomes for these coupling reactions. In the case of the Buchwald-Hartwig cross coupling the choice of a non-electron rich sulfonate is in agreement with published results. Further studies using different substrates might prove successful. In the pursue of different strategies to solve this problem design of experiment was carried out unsuccessfully. One suggestion for the improvement of design of experiment tools would be the conjugation of this statistical methodology with neural networks, thus lowering even more the value of "independent" variables taking into account variables "interactions" in a less iterative manner. Nevertheless, it is obvious to state that we were unable to find the best way out of this maze and suggestions may prove to be just that.



Scheme 4.6 – Alkylative cleavage of supported arylsulfonate esters using the Wustrow linker.[20]

Another possibility of introducing diversity at the final cleavage step that was investigated is the use of alkylative displacement reactions. It is expectable that our tetrafluoroarylsulfonate should display a similar or even higher reactivity than the Wustrow linker[20], thus increasing the usefull synthethic range of possible transformations with this linker (**scheme 4.6)**. Nevertheless the use of fluoride sources proved to be trigging since the product proved to keep the alcohol function instead of the expected fluorinated product (**scheme 4.5**).

4.9. References

(1) Fielding, H. C.; Shirley, I. M. *Journal of Fluorine Chemistry* **1992**, *59*, 15-31.

(2) Revell, J. D. PhD, Southampton, 2003.

(3) Pan, Y. J.; Holmes, C. P. *Organic Letters* **2001**, *3*, 2769-2771.

(4) Revell, J. D.; Ganesan, A. *Chemical Communications* **2004**, 1916-1917.

(5) Umetrics In *MODDE*; 7 ed.; Umetrics: Umea-Sweden, 2003; Vol. 1, p Software for Design of Experiments and Optimization.

(6) McNamara, C. A.; King, F.; Bradley, M. *Tetrahedron Letters* **2004**, *45*, 8527-8529.

(7) Carlson, R. C., J. *Design and optimization in organic synthesis*; ELSEVIER: Amsterdam, 2005.

(8) Owen, M. R.; Luscombe, C.; Lai, L. W.; Godbert, S.; Crookes, D. L.; Emiabata-Smith, D. *Organic Process Research & Development* **2001**, *5*, 308-323.

(9) Ballistreri, F. P.; Fortuna, C. G.; Musumarra, G.; Pavone, D.; Scire, S. *Arkivoc* **2002**, 54-64.

(10) Sarbu, C. *Revista De Chimie* **1999**, *50*, 852-863.

(11) Chaudhari, A.; Chaudhari, H.; Mehrotra, S. *Journal Of The Chinese Chemical Society* **2002**, *49*, 489-494.

(12) Chen, X. G.; Xu, Y. J.; Hu, Z. H. *Analytica Chimica Acta* **1995**, *314*, 213-218.

(13) Lou, J. F.; Hatton, T. A.; Laibinis, P. E. *Journal Of Physical Chemistry A* **1997**, *101*, 5262-5268.

(14) Cremin, P. A.; Zeng, L. *Analytical Chemistry* **2002**, *74*, 5492-5500.

(15) Fang, L. L.; Wan, M.; Pennacchio, M.; Pan, J. M. *Journal Of Combinatorial Chemistry* **2000**, *2*, 254-257.

(16) Hsu, B. H.; Orton, E.; Tang, S. Y.; Carlton, R. A. *Journal Of Chromatography B-Analytical Technologies In The Biomedical And Life Sciences* **1999**, *725*, 103-112.

(17) Mourey, T. H.; Oppenheimer, L. E. *Analytical Chemistry* **1984**, *56*, 2427-2434.

(18) Charlesworth, J. M. *Analytical Chemistry* **1978**, *50*, 1414-1420.

(19) Shekhar, S.; Ryberg, P.; Hartwig, J. F.; Mathew, J. S.; Blackmond, D. G.; Strieter, E. R.; Buchwald, S. L. *Journal Of The American Chemical Society* **2006**, *128*, 3584-3591.

(20) Jin, S. J.; Holub, D. P.; Wustrow, D. J. *Tetrahedron Letters* **1998**, *39*, 3651-3654.

Chapter Five

Experimental

5.1. General Section

Commercially available reagents were used without further purification. Solvents were not dried or distilled except where specified.

All solution-phase reactions were stirred magnetically and analysed by RP-HPLC or thin-layer chromatography (TLC) where appropriate using aluminium-coated Silica Gel 60 (Macheray Nagel: 0.20 mm layer). TLC visualisation was performed using short wavelength UV light (254 nm) and/or PMA oxidation.

NMR spectra were recorded on a Bruker DPX 400 and 300, or ARX 250 spectrometers in the solvents indicated at 298 K. Chemical shifts are reported on the δ scale in ppm and are referenced to residual non-deuterated solvent resonances. Signals of ^{1}H Magic Angle Spinning-NMR spectra were assigned by comparison with the ^{1}H NMR spectra obtained with the solution-phase equivalent of the supported compound. All ^{13}C NMR assignments were supported with DEPT.

MALDI spectra were obtained with an Applied Biosystems Voyager DESTR MALDI-TOF spectrometer.

IR spectra were obtained on a Thermo Mattson Satellite FTIR spectrometer or a Bruker Tensor 27 Spectrometer, 16 scans, at a resolution of ± 4 cm^{-1}. The FTIR spectrometers were fitted with a Specac single reflection diamond ATR Golden Gate, and neat compounds were used for analysis. Frequencies are reported in cm^{-1} and only frequencies corresponding to significant functional groups are reported.

LC-Mass spectra were recorded either on a Waters ZMD single quadrupole MS, with a 2700 Autosampler and a 600 Pump, or an Agilent Technologies LC/MSD Series 1100 Quadrupole Mass Spectrometer (QMS), both with an electrospray ion

source. HRMS analyses were performed by the Mass Spectrometry Service of the University of Edinburgh, U.K. Elemental analysis were carried out by Medac Ltd, U.K.

Peptide synthesis was performed by solid-phase Fmoc-chemistry using an automated microwave peptide synthesizer (Liberty, from CEM – 100W) .

UV/VIS spectra were recorded on a Hewlett Packard HP8452A diode array spectrophotometer.

Oxygen plasma irradiation of glass slides (obtained from Fisher Scientific, Silane-Prep^{TM}) was performed with a Europlasma instrument. Spin coating was performed using a Model: P6708 (Speedlines Technologies, IN, USA).

Contact printing was performed using a Genetix QArray mini (Hampshire, UK) with 16 aQu solid pins. The stampings were carried out with a delay of 200 ms between stampings, and the number of stampings is reported for each type of slide.

A Microdrop MD-E-401 printer equipped with an AD-K-501 autodrop pipette (Inner nozzle diameter 70μm) and an MD-O-538-85 CCD camera was used with the printing chamber kept at 70±5% relative humidity using a V5100N Vicks ultrasonic humidifier controlled by a humidity controller TH-810H (Advanced Timer Technologies LTD).

Microarrays were scanned with a CCD based fluorescence scanner (Bioanalyser 4F, LaVision Biotech) using fluorescein and Cy3 filters. Analysis of the microarray images was made using FIPS software (LaVision Biotech).

Image capture and analyses of microarray microwells slides was carried out using a Bioanalyzer 4F/4S fluorescent scanner (Lavision BioTec) or a Nikon Eclipse 50i microscope with the Pathfinder software (IMSTAR S.A., Paris, France).

All *p*H's were metered using a HANNA instruments HI 8424 microcomputer.

HPLC data were recorded using an Agilent 1100 Series coupled to a Polymer Lab 100 ES Evaporative Light Scattering Detector (ELSD), with a Phenomenex Luna C18, 5 μm, 10cm column (column 1), a Phenomenex Gemini C18, 5 μm, 10cm column (column 2), or a Phenomenex Luna C18, 5 μm, 15cm column (column 3). HPLC grade water, MeOH or CH_3CN with 0.1 % formic acid were used as eluent, at a flow rate of 1 mL/min., with samples prepared at a concentration of about 30

μg.mL^{-1} (ZMD) or 1 mg.mL^{-1} (Agilent) and filtered prior to injection. The following methods were used:

Method A (column 1, H_2O/MeOH, 12 min.): 0 min. (95/5), 7 min. (5/95), 9 min. (5/95).
Method B (column 2, H_2O/MeOH, 10 min.): 0 min. (95/5), 7 min. (5/95), 9 min. (5/95).
Method C (column 3, H_2O/MeOH, 15 min.): 0 min. (95/5), 12 min. (5/95), 14 min. (5/95).
Method D (column 1, H_2O/CH_3CN, 6 min.): 0 min. (95/5), 3 min. (5/95), 4 min. (5/95).

Preparative RP-HPLC was performed on a Hewlett Packard HP1100 Chemstation on a Waters (XTerra) RP18 OBD preparative column, 19 mm x 150 mm i.d. 5 μm at a flow rate of 10 ml min^{-1}, using the method below. The fractions were collected using a Gilson G2250A robot. The following method was used:

Method E: (A) 0.1% TFA/H_2O, (B) 0.1% TFA/CH_3CN; gradient from 100% (A) to 60% (A) in (B) over 3 min, 60 % (A) in (B) to 10 % (A) in (B) over 10 min, 10 % (A) in (B) to 0% (A) in 1 min, 0 % (A) to 100 % in 1 min, detection by UV at 214, 220, 254, 280, 440 and 650 nm.

5.1.1. General solid-phase chemistry procedures

Solid phase reactions were carried out in polypropylene syringes equipped with polyethylene frits and Teflon stopcocks.

5.1.1.1. Ninhydrin Analysis[1].

Reagent A

- Solution 1: Reagent grade phenol (40 g, 0.43 mol) was dissolved in absolute ethanol (10 mL). Amberlite mixed bed resin MB3 (4 g) was added, stirred for 45 min and afterwards filtered.

- Solution 2: KCN (65 mg, 1 mmol) was dissolved in water (100 mL). The KCN solution (2 mL) was diluted to 100 mL with pyridine (freshly distilled from ninhydrin). Amberlite mixed bed resin MB3 (4 g) was added, stirred for 45 min, filtered and mixed with Solution 2 to give reagent A.

Reagent B

- Ninhydrin (2.5 g, 14 mmol) was dissolved in absolute ethanol (50 mL).

Qualitative Test:

Reagent A (3 drops) and reagent B (1 drop) was added to a small sample of resin (< 0.5 mg) in a small test tube. The mixture was heated at 100 °C for 5 min. The intensity of blue colour gives a qualitative indication of the presence of the free primary amine groups.

Quantitative Test:

To an exact amount of dry resin (2-5 mg, W) in a small test tube, reagent A (6 drops) and reagent B (2 drops) was added. The mixture was heated at 100°C for 5 min. The tube was placed in a cold water bath and 40 % water in ethanol (2 mL) was added and mixed thoroughly. The solution was filtered and washed with 0.5 M Et_4NCl in CH_2Cl_2 (2 × 0.5 mL). The volume was made up to 50 mL (V) with 40 % water in

ethanol. The absorbance (A) of this solution was recorded at 570 nm against a control solution. The control solution was prepared exactly the same as a sample except without a resin added (each sample was prepared in duplicate):

Amount of amine present (mmol/g) = $[(A_{570} \times V\ (mL) \times 10^3) / (\varepsilon_{570} \times W\ (mg))]$

where ε_{570} is an average extinction coefficient with the value of 1.5×10^4.

5.1.1.2. Quantitative Fmoc Test[2].

A dry resin sample (ca 5 mg, W (mg)) was added to a 25 mL volumetric flask. A solution of 20 % piperidine in DMF (5 mL) was added and the mixture shaken for 15 min. The volume was made up to 25 mL with 20 % piperidine in DMF and the absorbance measured at 302 nm against a 20 % piperidine in DMF (value of ε for fulvene adduct is 7800).

$$\text{Loading of Fmoc (mmol/g)} = (A_{302} \times 25 \times 10^3)/(\varepsilon_{302} \times W\ (mg))$$

5.1.1.3. Calculation of theoretical loading.

The theoretical loading of a resin after a reaction was calculated using the equation:

$$\text{New Loading (mmol/g)} = [\text{Old loading}/(1+(\text{Old loading} \times \text{Mass added} \times 10^{-3})]$$

5.1.1.4. Activation of acidic ion exchange resin (Amberlite 200).

After washing Amberlite 200 (10 g) with ethanol/methanol (1:1 v/v, 3× 100 mL), the resin was rinsed with water (4 × 100 mL), washed with 1N aqueous hydrochloric acid (1× 100 mL), and rinsed again with water (2 × 100 mL). It was then washed with 1N aqueous NaOH (1 x 100 mL), rinsed with water (2 × 100 mL), and the acid form was generated by swelling the resin in concentrated aqueous hydrochloric acid

in water (37% v/v, 2 × 100 mL). The beads were then washed with methanol (5 × 100 mL) to remove water.

5.1.2. Buffers for Protease substrate specificity assays.

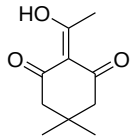
5.1.2.1. Buffers for microarray printing and washing.

- ***6X Print buffer*** [300mM sodium phosphate (pH 8.5)]. Sodium phosphate monobasic (0.41 g) and sodium phosphate dibasic (3.785 g) were dissolved in 90 mL of nuclease-free distilled water. The pH was adjusted to 8.5 using 1N NaOH or 1N HCl, and the volume was brought to 100 mL with nuclease-free distilled water.
- ***Blocking Solution*** [0.1 M Tris, 50 mM ethanolamine (pH 9.0)]. Trizma Base (6.055 g), Trizma HCl (7.88 g), and ethanolamine (3.05 g, 3.0 mL) were dissolved, with stirring, in 900 mL of nuclease-free distilled water. The pH was adjusted to 9.0 using 6N HCl, and the volume was brought to 1 L with nuclease-free distilled water.
- ***10 % SDS***. Sodium dodecyl sulphate (100 g) was dissolved into 900 mL of nuclease-free distilled water, and the solution was heated slightly to solubilise the solid. The pH was adjusted to 7.2 by adding a few drops of 6N HCl, and the volume was brought to 1 L with nuclease-free distilled water.
- ***20X SSC***. Sodium chloride (175.3 g) and sodium citrate (88.2 g) were dissolved into 800 mL of nuclease-free distilled water. The pH was adjusted to 7.0 with a few drops of 10N NaOH and the volume was brought to 1 L with nuclease-free distilled water.

5.2. Experimental for Chapter Two.

5.2.1. Preparation of the dual colour FRET peptide library.

- **1-hydroxy-1-(4,4-dimethyl-2,6-dioxacyclohexyldene)ethane[3] 2.5.**



$C_{10}H_{14}O_3$

Mol. Wt.: 182.09

To a mixture of DCC (2.95 g, 14.3 mmol, 1eq) and DMAP (1.74 g, 14.3 mmol) in acetic acid (0.82 mL, 14.3 mmol), dimedone (2.0 g, 14.3 mmol) dissolved in DMF (35 mL) was added at room temperature. The mixture was heated at 60°C. After 16h, more DCC (0.295g 1.43 mmol) and acetic acid (0.082 mL) were added. The reaction was followed by analytical TLC (CH_2Cl_2/CH_3OH 95:5 v/v) and was complete over 36h (t_R = 0.97). Precipitated *N,N'*-dicyclohexylurea (DCU) was removed by filtration. The solvent was evaporated *in vacuo* and the residue was dissolved in ethyl acetate (35 mL). The remaining traces of DCU were removed by filtration through a Celite plug and the organic layer was washed with 1M aqueous $KHSO_4$ (3 x 30 mL) and then dried with $MgSO_4$. The solvent was removed *in vacuo* to afford 1-hydroxy-1-(4,4-dimethyl-2,6-dioxacyclohexyldene)ethane as an orange oil (2.4 g; 92 % yield).

HPLC (method 2): t_R = 9.8 min.

Purity**:** 98% (UV 254 nm)

ESI+/MS**:** m/z 183.2 $(M+H)^+$

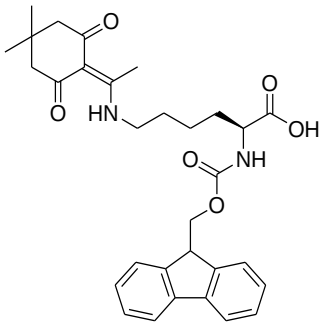
IR ν (cm^{-1}): 3307 (vw, br, ν_{O-H}), 1659 (vs, $\nu_{C=O}$), 1543 (s, $\nu_{C=O}$)

^{1}H NMR (300MHz, $CDCl_3$, δ, ppm): 2.59 (3H, s), 2.52 (2H, s), 2.34 (2H, s), 1.06 (6H, s).

^{13}C NMR + DEPT 135 + DEPT 90 (75 MHz, $CDCl_3$, δ, ppm): 202.3, 197.8, 195.1, 112.3, 52.3, 47.1, 30.7, 28.6, 28.3.

These data were similar to the literature values.[3]

- **Fmoc-Lys(Dde)-OH[4] 2.6.**



$C_{31}H_{36}N_2O_4$

Mol. Wt.: 532.63

Trifluoroacetic acid (105 μL, 1.35 mmol) was added to a stirred suspension of Fmoc-Lys-OH (5 g, 13.5 mmol) and Dde-OH (5g, 27 mmol) in ethanol (112.5 mL) at room temperature. The mixture was refluxed for 60 hours, followed by analytical TLC (ethyl acetate/hexane 95:5 v/v; R_f = 0.70). The solvent was removed *in vacuo* and the orange residue dissolved in ethyl acetate (150 mL). The organic solution was washed with 1M aqueous $KHSO_4$ (2 x 100 mL), dried with $MgSO_4$, filtered and the solvent was evaporated until a yellow oil was obtained. The yellow oil was triturated with hexane to remove unreacted Dde-OH. The product was precipitated from CH_2Cl_2/hexane, filtered and dried *in vacuo* to afford the desired product as a white powder (5.8 g; 80 % yield).

Purity: 98% (UV ε coeficiente at 254 nm)

ESI+/MS: m/z 533.4 ($M+H^+$)

IR ν (cm^{-1}): 3304 (w, br, ν_{N-H}), 1715 (vs, $\nu_{C=O}$), 1630 (m, $\nu_{C=O}$), 1534 (vs, br, $\nu_{C=O}$)

^{1}H NMR (300 MHz, $CDCl_3$, δ, ppm): 13.43 (1H, s, COOH), 9.13 (1H, bs, C(O)NH), 7.73 (2H, d, J^3 = 7.1 Hz, H_{ar}), 7.56 (2H, d, J^3 = 6.6 Hz, Har), 7.39 (2H, t, J^3 = 6.7 Hz, H_{ar}), 7.28 (2H, d, J^3 = 7.4 Hz, H_{ar}), 5.85 (1H, bs, J^3= 8.0 Hz, CH_2NH), 4.43 (1H, m, CH), 4.35 (2H, d, J^3= 6.8 Hz, CH_2 Fmoc), 4.18 (1H, t, J^3 = 6.7 Hz, CH Fmoc), 3.37 (2H, m, CH_2NH), 2.52 (3H, s, CH_3 Dde), 2.35 (4H, s, 2 × CH_2 Dde), 1.96 (2H, m, CH_2 lys) 1.72 (2H, m, CH_2 lys), 1.51 (2H, m, CH_2 lys), 1.00 (6H, s, $(CH_3)_2$ Dde)

^{13}C NMR + DEPT 135 + DEPT 90 (75 MHz, $CDCl_3$, δ, ppm): 198.6 (C(O) Dde), 174.8 (C(O)OH), 174.4 (C=CNH), 156.4 (C(O)NH), 144.1 (C_{ar} Fmoc), 141.5 (C_{ar} Fmoc), 128.0 (CH_{ar} Fmoc), 127.3 (CH_{ar} Fmoc), 125.4 (CH_{ar} Fmoc), 120.3 (CH_{ar} Fmoc), 108.1 (C=CNH), 67.3 (CH_2 Fmoc), 53.6 (HOOC-CH), 52.5 (2 × CH_2 Dde), 47.3 (CH Fmoc), 43.5 (CH_2NH), 32.2 (CH_2 lys), 31.7 (CH_2 lys), 30.2 ($C(CH_3)_2$), 28.7 ((C(CH_3)$_2$)), 22.8 (CH_2 lys), 18.3 (CH_3 Dde).

These data were similar to those reported in the literature.[3]

- Pre-activation of 2-chlorotrityl resin.

2-Chlorotrytil resin (2 g; 1.2 mmolg^{-1}) was swollen in CH_2Cl_2 (15 mL) for 5 minutes and activated by stirring with $SOCl_2$ (0.32 mL; 2 eq) and Et_3N (0.612 mL; 2 eq) for 2 hours. The resin was then washed with DMF (3 x 15 mL), CH_2Cl_2 (3 x 15 mL) and dried overnight at 40^{o}C under reduced pressure and over phosphorus pentoxide.

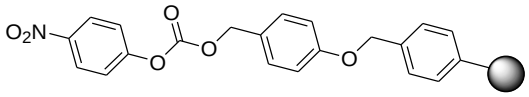
- Attachment of **2.11** to 2-chlorotrityl resin.

To a batch of 2-chlorotrytil resin (2 g; 1.2 mmolg^{-1}) swollen in CH_2Cl_2/DMF (20 ml of a 3:1 v/v mixture), was added 1,2-bis(2-amino ethoxy)ethane **2.11** (3.22 mL; 10 eq.) and Et_3N (3.06 mL; 10 eq.). The resin was stirred for 4 hours, filtered and washed with DMF (3 x 20 mL), CH_2Cl_2 (3 x 20 mL), CH_3OH (3 x 20 mL) and Et_2O (3 x 20 mL). The resin was dried overnight at 40^{o}C under reduced pressure. Final loading was determined through a quantitative ninhydrin test (0.82 mmolg^{-1}).

- Attachment of **2.6** to 2-chlorotrityl resin.

Fmoc-Lys(Dde)-OH (**2.6**) (2.5g; 4.63 mmol; 1.5 eq.) was dissolved in CH_2Cl_2/DMF (7:3 v/v, 25mL), HOBt was then added (0.624 g; 4.63 mmol; 1.5 eq). After complete dissolution, DIC (721 μL; 4.63 mmol; 1.5 eq.) was added. The mixture was stirred for 5 minutes and added to the resin (2g; 1eq). The reaction was stirred for 2 hours, time when a qualitative ninhydrin test was negative. The resin was washed with DMF (3 x 20 mL), CH_2Cl_2 (3 x 20 mL), CH_3OH (3 x 20 mL) and Et_2O (3 x 20 mL). The resin was dried overnight at 40^oC under reduced pressure to afford a clear yellow resin. The removal of the Fmoc group was performed by swelling the resin in a solution of piperidine in DMF (20 % v/v) for 2 cycles of 15 minutes. Between cycles the resin was washed with DMF (3 x 20 mL) and CH_2Cl_2 (3 x 20 mL). Afterwards the resin was washed with DMF (3 x 20 mL), CH_2Cl_2 (3 x 20 mL), CH_3OH (3 x 20 mL) and Et_2O (3 x 20 mL), and dried overnight at 40^oC under reduced pressure to afford the final product.

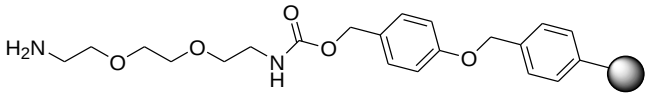
- Preparation of Wang Carbonate resin **2.9**.



Polystyrene resin functionalised with a Wang linker (1 g; 0.92 mmolg^{-1}) was swollen in CH_2Cl_2 (10 mL) and pyridine (0.446 mL; 5.52 mmol; 6 eq.) for 10 minutes. 4-Nitrophenyl chloroformate **2.8** (0.556 g; 2.76 mmol; 3 eq.) was added and the mixture was stirred for 4 hours, washed with DMF (10 ml; 3x) and DCM (10 ml; 3x) and dried at 40^{o}C under vacuum.

IR ν (cm^{-1}): 1764 (s, br, $\nu_{C=O\ ester}$).

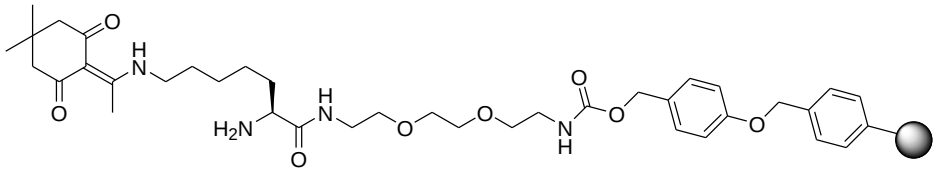
- Attachment of **2.11** to Wang carbonate resin.



The previously prepared Wang carbonate resin **2.9** was pre-swollen in DCM (10 mL per g of resin) for 10 minutes, after which the supernatant was removed and 1,2-bis(2-aminoethoxy)ethane was added (10 mL per g of resin). The resin was stirred for 4 hours, washed with DMF (10mL g^{-1}; 3x) and DCM (10mL g^{-1}; 3x), CH_3OH (10 ml, 3x), diethyl ether (10 ml, 3x) and dried at 40^{o}C under vacuum.

IR ν (cm^{-1}): 1718 (s, br, $\nu_{C=O\ amide}$).

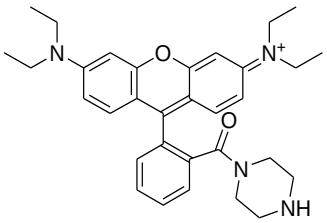
- Attachment of **2.6** to **2.12.**



Fmoc-Lys(Dde)-OH (1.25g, 2.315 mmol) was dissolved in CH_2Cl_2/DMF (7:3 v/v, 12.5 mL), in the presence of HOBt (312 mg, 2.315 mmol). After complete dissolution, DIC (360 μL, 2.315 mmol) was added; the mixture was shaken for 5 minutes and then added to the resin (1g, loading=0.82 $mmolg^{-1}$). The resin was shaken until qualitative ninhydrin tests were negative (approximately 2 hours). The resin was washed with DMF (10 ml, 3x), CH_2Cl_2 (10 ml, 3x), CH_3OH (10 ml, 3x) and diethyl ether (10 ml, 3x). The resin was dried overnight *in vacuo* at 40^oC to afford a clear yellow resin with a loading of 0.68 $mmolg^{-1}$ (determined by quantitative Fmoc test). The deprotection of the Fmoc group was performed by swelling the resin in a solution of piperidine in DMF (20% v/v; 10 ml, 2x). The resin was then washed with DMF (10 ml, 3x), CH_2Cl_2 (10 ml, 3x), CH_3OH (10 ml, 3x) and diethyl ether (10 ml, 3x). The resin was dried overnight *in vacuo* at 40 ^oC to afford a clear yellow resin.

IR ν (cm^{-1}): 3327 (w, br, ν_{N-H}), 1676 (s, $\nu_{C=O}$), 1644 (m, $\nu_{C=O}$), 1603 (m, $\nu_{C=O}$), 1575 (m, $\nu_{C=O}$).

- Synthesis of Rhodamine B piperazine amide **2.20**[5].



$C_{32}H_{39}N_4O_2^+$

Mol. Wt.: 511.68

A 2.0 M solution of trimethyl aluminium in toluene (22.6 mL, 45.2 mmol) was added dropwise to a solution of piperazine (7.8 g, 91 mmol) in 35 mL of DCM at room temperature. After one hour of stirring a white precipitate was observed. A solution of rhodamine B base (10.0 g, 22.6 mmol) in 20 mL of DCM was added dropwise to the heterogeneous solution. Gas evolution was observed during the addition period. After stirring at reflux for 24 h, a 0.1 M aqueous solution of HCl was added dropwise until gas evolution ceased. The heterogeneous solution was filtered and the retained solids were rinsed with DCM and a 4:1 DCM/MeOH solution. The combined filtrate was concentrated and the residue was dissolved in DCM, filtered to remove insoluble salts, and concentrated again. The resulting glassy solid was then partitioned between dilute aqueous $NaHCO_3$ and EtOAc. After isolation, the aqueous layer was washed with 3 additional portions of EtOAc to remove residual starting material. The retained aqueous layer was saturated with NaCl, acidified with 1 M aqueous HCl, and then extracted with multiple portions of 2:1 *i*PrOH/DCM, until a faint pink colour persisted. The combined organic layers were then dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The glassy purple solid was dissolved in a minimal amount of MeOH and precipitated by dropwise addition to a large volume of Et_2O. The product was collected by filtration as a dark purple solid 8.7 g, (70% yield).

M.p. 218-220^{o}C.

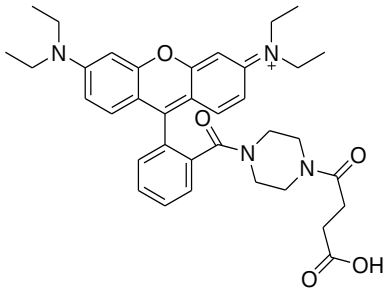
IR: 1630 cm^{-1}.

^{1}H NMR (300 MHz, CD_3OD): δ 1.27-1.33 (t, 12, *J*=7.4), 3.12 (br s, 4), 3.64-3.74 (m, 12), 6.96-6.98 (d, 2, *J*=2.4), 7.09-7.12 (dd, 2, *J*=2.4, 10.0), 7.26-7.28 (d, 2, *J*=9.2), 7.50-7.54 (m, 1), 7.75-7.80 (m, 3).

^{13}C NMR (75 MHz, CD_3OD) δ 13.06, 44.38, 45.64, 47.01, 97.50, 114.84, 115.68, 129.02, 131.51, 131.66, 132.00, 132.55, 133.05, 135.77, 156.85, 157.32, 159.34, 169.51.

These data were similar to those reported in the literature.[5]

- Rhodamine B 4-(3-Carboxypropionyl)piperazine amide[5] **2.21**.



$C_{36}H_{43}N_4O_5^+$

Mol. Wt.: 611.75

Triethylamine (0.05 mL, 0.4 mmol) was added to a stirred solution of **2.20** (0.2 g, 0.3 mmol), succinic anhydride (0.04 g, 0.4 mmol), and DMAP (0.05 g, 0.4 mmol) in DCM (1.2 mL). After stirring at room temperature for 24 h, the reaction solution was partitioned between EtOAc and 1 M aqueous K_2CO_3. The aqueous layer was washed with 3 additional portions of EtOAc. Sodium chloride was added to the isolated aqueous layer until saturation was achieved and the solution was then extracted with 2:1 isopropanol/DCM. The organic layer was dried over Na_2SO_4, and concentrated under reduced pressure. The resulting solid was dissolved in $CHCl_3$ and filtered to remove insoluble salts. Upon concentration, the product was obtained as a dark solid, 144 mg (72 % yield).

Mp: 167-168^{o}C (crystalized from $CHCl_3$).

IR: 1720 cm^{-1}.

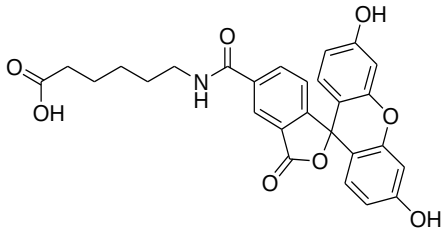
^{1}H NMR (300 MHz, CD_3OD) δ 1.27-1.31 (t, 12H, *J*=6.8), 2.54 (br s, 2H), 2.60 (br s, 2H), 3.40 (br s, 8H), 3.65-3.71 (q, 8H, *J*=7.2), 6.95-6.97 (d, 2H, *J*=2.0), 7.09 (br s, 2H), 7.26-7.29 (d, 2H, *J*=9.6), 7.51-7.52 (m, 1H), 7.69-7.71 (m, 1H), 7.75-7.77 (m, 2H).

^{13}C NMR (750 MHz, CD_3OD) δ 13.0, 28.97, 30.40, 40.45, 43.14, 47.08, 50.04, 97.52, 114.97, 115.60, 129.1, 131.45, 131.92, 132.42, 133.31, 136.68, 140.17, 157.14, 157.33, 169.67, 172.84, 176.72.

MS, m/z (M^+) 611.32.

These data were similar to those reported in the literature[5].

- Derivative of 5(6) carboxyfluorescence dye **2.25**.



$C_{27}H_{23}NO_8$

Mol. Wt.: 489.47

A solution of 5(6)-carboxyflourescein **2.22** (5g, 13 mmol) in 12.5 mL of DMF was slowly added to a solution of **2.23** with HBTU (1.05 eq. to **2.22,** 15 mmol) in 12.5 mL of DMF. The solution was stirred under reflux, for 2 hours. The solvent was evaporated and the remaining slurry was dissolved in 20 mL of a 1M solution of NaOH (1:1 H_2O/THF) and stirred for 30 minutes. Afterwards the solution was washed with a 2:1 *i*PrOH/DCM solution (3 x), and the organic layers were collected, dried with Na_2SO_4 and evaporated to afford the final product as a bright yellow powder (95% yield after purification).

IR: 1722 cm^{-1}.

^{13}C NMR (100 MHz, CD_3OD) δ 24.00, 26.31, 30.33, 36.40, 41.08, 93.05, 105.80, 110.97, 113.18, 127.33, 128.00, 128.40, 129.07, 131.42, 132.90, 152.61, 155.29, 155.77, 167.54, 169.52, 177.32.

MS, m/z (M+) 490.51.

- General automated microwave peptide coupling procedure.

Automated peptide synthesis was performed by solid-phase Fmoc/tBu-chemistry using an automated peptide synthesizer for multiple peptide synthesis (CEM Liberty Microwave Assisted Peptide) in using a 5 mL reaction chamber. Fmoc-amino acids (in a 5 eq. relation to the resin load) were coupled using single couplings of HBTU/HOBt (5 eq.) and DIPEA/NMP (1:4 v/v, 10 eq. of DIPEA) under microwave irradiation for 20 minutes at a temperature of 60^{o}C. The Fmoc protecting group was removed by treatment with piperidine/DMF (1:4 v/v) twice for 8 min. The resin was washed with DMF (3 x 7 mL) and DCM (3 x 7 mL) after each coupling and deprotection step.

- General manual Fmoc deprotection procedure.

To the pre-swollen 100 mg of each batch of resin (swollen in DCM) was added a solution of 20% piperidine in DMF (1 mL) and the resins were shaken for 10 minutes. The resins were washed with DMF (3 x 1 ml) and DCM (3 x 1 ml). The resins were allowed to stand dry for 5minutes and the 20% piperidine solution in DMF (1 ml) was added and the resins shaken for 10 minutes once again. Afterwards, the resin was washed with DMF (3 x 2 mL), CH_2Cl_2 (3 x 2 mL), CH_3OH (3 x 2 mL) and diethyl ether (3 x 2 mL). The resins were dried overnight *in vacuo* at 40^{o}C. A qualitative ninhydrin test was performed prior and afterwards to verify the deprotection of the Fmoc group.

- General coupling procedure for 4-dimethylaminoazobenzene-2-carboxylic acid.

A solution of 4-Dimethylaminoazobenzene-2′-carboxylic acid (5eq to the resin loading), HBTU(5eq to the resin loading), HOBt (5eq to the resin loading) in DMF (1ml per gram of resin) was added to the resin. The resin was shaken overnight and then washed with DMF (5 x 1 ml per gram of resin), DCM (5 x 1ml per gram of resin) and MeOH (5 x 1 ml per gram of resin). The coupling procedure was repeated

once more overnight and the resin was then washed with DMF (5 x 2 ml per gram of resin), CH_2Cl_2 (5 x 2 ml per gram of resin), CH_3OH (5 x 2 ml per gram of resin) and diethyl ether (5 x 2 ml per gram of resin) and dried overnight *in vacuo* at 40^{o}C. Qualitative ninhydrin tests were performed after the first overnight coupling and proved the coupling was incomplete. After the second coupling cycle the coupling was complete.

Note: Extending the reaction time to attempt a single coupling proved to lead to "dirty" resins in the sense that normal washing cycles were not enough to stop the resin from leaching a slight red colour after several (> 10 x) washing cycles.

- General Dde deprotection procedure.

Resins containing a general peptide sequence (after attachment of the quencher dye) were treated twice using 2% hydrazine hydrate in DMF for 3 min and washed thoroughly with DMF (3x 5 mL) and DCM (3x 5 mL).

- Fluorophores **2.21** and **2.25** coupling procedure.

In all experiments, the isomeric mixture of 5-and 6-carboxyfluorescein (61:39 isomer ratio, purchased from Fluka) was used for labelling, unless otherwise stated. The labelling using **2.25** was performed by in situ activation using 2.5 equiv of dye, DIC, and HOBt (1:1:1) for 5 minutes, addition to the resin and overnight coupling (10 hours). Treatment of the resin(s) with piperidine/DMF (1:4, v/v) removed side products, yielding highly pure products (sometimes a second coupling cycle was required). The resin was afterwards washed with DMF (3x 5mL) and DCM (3x 5 mL).

The labelling using **2.21** was performed by in situ activation using 2.5 equiv of dye, DIC, and HOBt (1:1:1) for 5 minutes, addition to the resin and overnight coupling (16 hours). Treatment of the resin with 2% hydrazine hydrate in DMF for 15 min removed remaining dye inside the resin (sometimes a second coupling cycle

was required). The resin was afterwards washed with DMF (3x 5mL) and DCM (3x 5 mL).

- Peptide cleavage procedure

A solution containing 47.5% TFA in 47.5% DCM and 5% TIS was added to the resin in a ratio of 10ml of solution per gram of resin. The resin was stirred for 30 minutes and the mixture was filtered and collected. The cleavage procedure was repeated three times to ensure maximum release of compound. The resin was then washed (on average) three times with neat DCM. On average three cleavages were found to ensure maximum release of compound from the resin. However, sometimes the presence of some colour in the final washing solution was found and an extended number of cleavage cycles were sometimes necessary.

- Library purification and characterization

Each compound of the library was purified by preparative RP-HPLC (Method E). Every member of the library was lyophilised after purification, characterized by MALDI-TOF and purity determined by ELSD and UV using method D.

Table 5.1 - HPLC and MALDI-TOF data on sublibrary subtilisin (R stands for rhodamine derivative and, F stands for FAM derivative)

Entry	Sequence	Rt/min Prep	Rt/min HPLC	Purity		MALDI-TOF	
				ELSD	UV	Expected	Obtained*
1	VNQHLK-R	7.469	4.715	99.9	99.9	1711.948	1713.644
2	EALYLK-R	9.177	5.440	99.9	99.9	1709.946	1712.240
3	FFYTPK-R	9.693	6.062	99.9	99.9	1776.939	1778.145
4	ALYLVK-R	10.338	6.091	99.9	99.9	1679.972	1682.249
5	QHLCGK-R	7.059	5.022	99.9	99.9	1658.867	1661.864
6	LCGSHK-R	7.809	5.012	99.9	99.9	1617.84	1620.782
7	VNQHLK-F	7.025	4.824	90	80	1589.767	1592.717
8	EALYLK-F	9.074	6.347	99.9	99.9	1587.765	1589.891
9	FFYTPK-F	9.838	5.778	99.9	99.9	1653.754	1655.487
10	ALYLVK-F	10.675	6.029	99.9	99.9	1557.791	1559.827
11	QHLCGK-F	6.895	4.660	99.9	99.9	1536.686	1539.311
12	LCGSHK-F	7.905	5.126	99.9	99.9	1495.660	1497.565

***** - Values presenting more than 1 Dalton divergency from expected were confirmed by ES^+ verifying (M+2)/z values.

Table 5.2 - HPLC and MALDI-TOF data on sublibrary thermolysin (R stands for rhodamine derivative and, F stands for FAM derivative)

Entry	Sequence	Rt/min Prep	Rt/min HPLC	Purity (%)		MALDI-TOF	
				ELSD	UV	Exact mass	Obtained*
1	YGGFL-R	8.996	5.744	99.9	99.9	1657.894	1658
2	GFFYY-R	9.488	5.595	99.9	99.9	1798.923	1799
3	EALYL-R	9.177	5.440	99.9	99.9	1709.946	1712.240
4	YQLEN-R	8.144	5.633	99.9	99.9	1767.926	1769.607
5	YNQFF-R	8.766	5.073	99.9	99.9	1820.940	1821.266
6	GSHLV-R	7.365	5.157	99.9	99.9	1613.900	1615.647
7	YGGFL-F	8.692	6.062	99.9	99.9	1535.713	1537.458
8	GFFYY-F	9.347	5.779	99.9	99.9	1675.739	1678.599
9	EALYL-F	9.074	6.347	99.9	99.9	1587.765	1589.891
10	YQLEN-F	7.684	5.907	99.9	99.9	1645.745	1648.841
11	YNQFF-F	8.342	5.367	99.9	99.9	1697.756	1700.957
12	GSHLV-F	6.935	4.640	99.9	99.9	1491.719	1492

***** - Values presenting more than 1 Dalton divergency from expected were confirmed by ES^+ verifying (M+2)/z values.

Table 5.3 - HPLC and MALDI-TOF data on sublibrary chymopapain (R stands for rhodamine derivative and, F stands for FAM derivative)

Entry	Sequence	Rt/min Prep	Rt/min HPLC	Purity (%)		MALDI-TOF	
				ELSD	UV	Expected	Obtained*
1	SALFI-R	10.052	5.875	99.9	99.9	1651.941	1653.505
2	ALFIA-R	10.542	5.335	99.9	99.9	1635.946	1637.847
3	SLFIA-R	9.872	5.892	99.9	99.9	1651.941	1653.504
4	SAFIA-R	9.359	5.500	99.9	99.9	1609.894	1611.640
5	SALIA-R	9.224	5.923	99.9	99.9	1575.909	1577.475
6	SALFL-R	10.261	5.693	99.9	99.9	1651.941	1654.014
7	SALFI-F	10.335	5.875	99.9	99.9	1529.760	1529.
8	ALFIA-F	10.918	6.296	99.9	99.9	1513.765	1515.745
9	SLFIA-F	10.318	6.145	99.9	99.9	1529.760	1531.512
10	SAFIA-F	9.568	5.935	99.9	99.9	1487.713	1489.245
11	SALIA-F	9.516	5.971	99.9	99.9	1453.728	1455.657
12	SALFL-F	10.549	6.142	99.9	99.9	1529.760	1529.947

***** - Values presenting more than 1 Dalton divergency from expected were confirmed by ES^+ verifying (M+2)/z values.

Table 5.4 - HPLC and MALDI-TOF data on sublibrary thermolysin (R stands for rhodamine derivative and, F stands for FAM derivative)

Entry	Sequence	Rt/min Prep	Rt/min HPLC	Purity		MALDI-TOF	
				ELSD	UV	Expected	Obtained*
1	PIQQH-R	7.261	5.266	99.9	99.9	1723.948	1726.094
2	HIPPY-R	7.547	5.086	99.9	99.9	1728.950	1729.730
3	PIHYQ-R	7.473	4.781	99.9	99.9	1759.956	1760.666
4	HPIQQ-R	7.065	4.985	99.9	99.9	1723.948	1725.639
5	IHPQP-R	8.594	5.734	99.9	99.9	1692.941	1694.937
6	PIPIS-R	8.642	5.956	99.9	99.9	1625.961	1629.826
7	PIQQH-F	6.967	4.624	95	90	1601.767	1602.978
8	HIPPY-F	7.232	4.928	90	80	1605.766	1608.796
9	PIHYQ-F	7.301	4.632	96	90	1636.772	1639.748
10	HPIQQ-F	6.797	4.399	99.9	99.9	1601.767	1603.657
11	IHPQP-F	7.637	4.496	92	90	1570.761	1572.681
12	PIPIS-F	8.622	5.528	99.9	99.9	1503.780	1503.945

***** - Values presenting more than 1 Dalton divergency from expected were confirmed by ES^+ verifying (M+2)/z values.

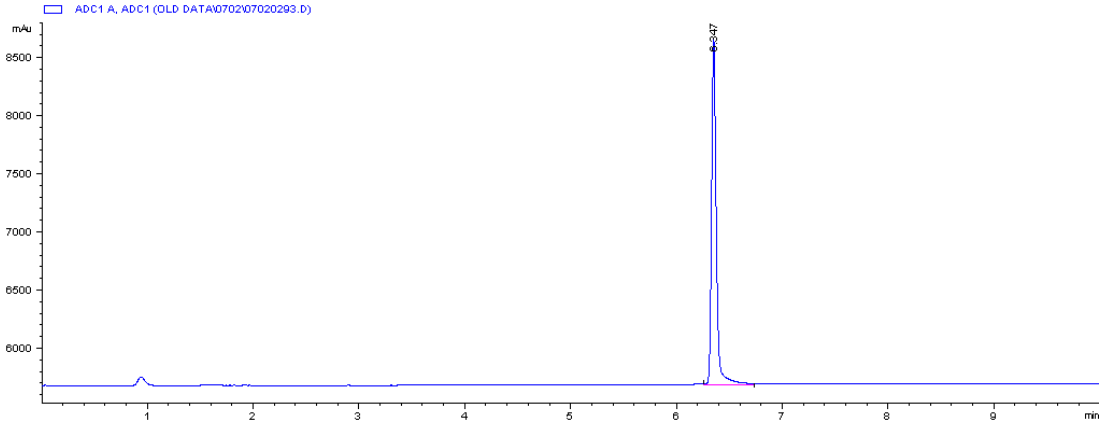


Figure 5.1 – Reprentative HPLC/ELSD trace of the final purified peptides (EALYLK-F).

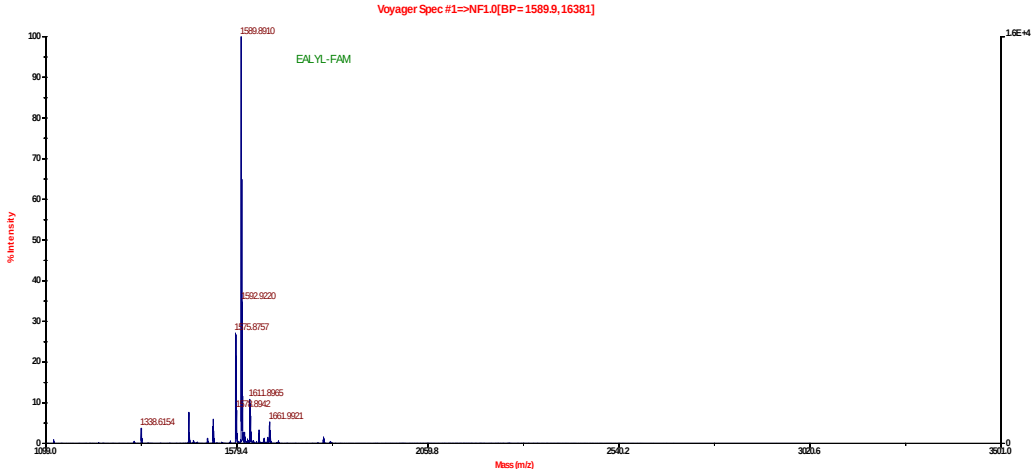


Figure 5.2 – Representative MALDI-TOF of the final purified peptideS (EALYLK-F)

.

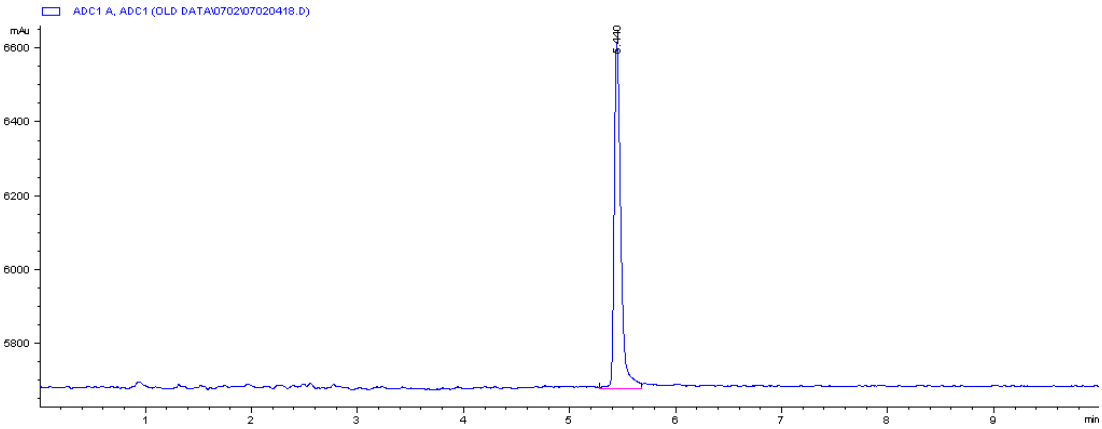


Figure 5.3 – Reprentative HPLC/ELSD trace of the final purified peptides (EALYLK-R).

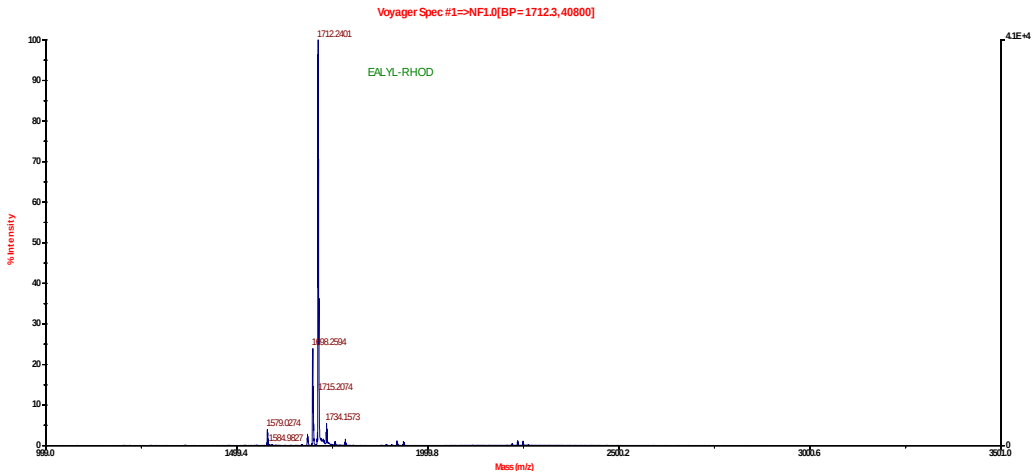


Figure 5.4 – Representative MALDI-TOF of the final purified peptideS (EALYLK-R).

5.2.2. Dual colour FRET peptide library assays.

5.2.2.1. Glycerol droplets and spray assay.

The library was always printed (by contact printing) in a 4 x4 format (16 spots per peptide; concentration of 25 μM). In different assays the constructs were printed in glycerol or in a gelator being developed in-house. A spray of water was used to "solvate" the printed droplets and afterwards the 250 μl of solution enzyme were sprayed (for units of enzymes see **table 5.5**). Subsequently the slides were incubated in a saturated chamber and scanned every hour. Some assays were attempted in a non saturated atmosphere but as expected resulted only in dry droplets with absolutely no signal. Attempts were carried out in order to analyse the influence of the nature of the glass surface without the attachment of the substrates. In this assay, the glass slides were coated with (tridecafluoro-1, 1, 2, 2-tetrafluorooctyl) dimethylchlorosilane in order to observe the influence of a hydrophobic surface. All the described assays data are presented in annex and in all slides the pattern used for the printing had the position pattern presented on **table 5.5** (A1 position being the one to the top right corner of the slide).

All the data treated in this assay was made using the software FIPS-2 from LaVision and is available in the support DVD with the data (the data for this section is available in DVD-1).

Table 5.5 – Order of the peptides on the plate from which were printed.

Sequences	Pos.	Sequences	Pos.	Sequences	Pos.	Sequences	Pos.
YGGFLK	A1	SALFIK	A2	VNQHLK	A3	PIQQHK	A4
GFFYYK	B1	ALFIAK	B2	EALYLK	B3	HIPPYK	B4
EALYLK	C1	SLFIAK	C2	FFYTPK	C3	PIHYQK	C4
YQLENK	D1	SAFIAK	D2	ALYLVK	D3	HPIQQK	D4
YNQFFK	E1	SALIAK	E2	QHLCGK	E3	IHPQPK	E4
GSHLVK	F1	SALFLK	F2	LCGSHK	F3	PIPISK	F4

- *Enzymes*

Table 5.6 – Enzyme conditions used in these assays

Enzyme	Units per *µ*l	Buffer
Subtilisin A	0.5	0.1 M Tris pH 8.0; 5 mM $CaCl_2$
Thermolysin	1	0.1 M Tris pH 8.0; 10 mM $CaCl_2$
Chymopapain	1	0.2 M NaCl, 6. mM Mercaptoethanol; 5 mM EDTA; 50 mM *L*-Cysteine hydrochloride hydrate

5.2.2.2. Assays on covalently linked peptides to glass slides.

- Printing and washing

  - *Attachment to CodeLink Slides*

CodeLink activated slides are coated with a hydrophilic polymer containing N-hydroxysuccinimide (NHS) ester reactive groups and designed to minimize non-specific binding. Amino containing compounds were coupled to the surface at pH=8.5 in a 75% relative humidity environment. Printing was carried out in a 50 mM sodium phosphate buffer solution (1x printing solution pH=8.5) and compounds were printed with a concentration of 25 µM. After printing the slides were placed

inside a slide storage box containing filter paper soaked in NaCl saturated solution. The box was sealed and incubation was performed at room temperature for a period of 36 hours (incubation can be performed for a minimum of 4 hours and a maximum of 72 hours). As a post-coupling process the remaining active sites of the slides were blocked using two different processes: 0.1 M Tris, 50 mM ethanolamine (pH=9); 0.1 M Tris, 50 mM using **2.11** (pH=9). The slides were placed in a slide rack and the pre-warmed blocking solution was added at 50^{o}C for 30 minutes. After, the blocking solution was discarded and the slides were individually rinsed twice with deionized water. The slides were then washed with a passivation/washing solution of 4x SSC, 0.1% SDS and 0.1% BSA (w/v), pre-warmed to 50^{o}C for 30 minutes in the shaker (using a minimum of 10 ml per slide). The passivation/washing solution was then discarded and the slides were individually rinsed twice with deionized water and carefully dried under a strong nitrogen flow, avoiding any type of staining. If the slides presented some staining a rewashing step was performed using 4x SSC and o.1% SDS at 50^{o}C for 30 minutes, followed by double rinsing and drying. The slides were then stored in a desiccated environment until usage.

- *Attachment to aldehyde functionalized slides*

Printing was carried out in a 50 mM sodium phosphate buffer solution (1x printing solution pH=8.5) and compounds were printed with a concentration of 25 μM. After printing the slides were placed inside a slide storage box containing filter paper soaked in NaCl saturated solution. The box was sealed and incubation was performed at room temperature for a period of 36 hours (incubation can be performed for a minimum of 4 hours and a maximum of 72 hours). As a post-coupling process the slides were washed twice with a 0.2% SDS solution for periods of 2 minutes, and then were rinsed twice with deionized water. The slides were then reduced using a sodium borohydride solution for 5 minutes (100 mg of sodium borohydride, 30 ml of 0.1M phosphate buffer, 10 ml of ethanol). The slides were then washed twice with a 0.2% SDS solution for periods of 2 minutes, and then were rinsed twice with deionized water and carefully dried under a strong nitrogen flow, avoiding any type of staining. If the slides presented some staining a rewashing step was performed

using 0.2% SDS at 50°C for 30 minutes, followed by double rinsing and drying. The slides were then stored in a desiccated environment until usage.

- *Attachment to AmpliSlide E Slides*

Printing was carried out in a 50 mM sodium phosphate buffer solution (1x printing solution pH=8.5) and compounds were printed with a concentration of 10 μM, under controlled environmental conditions of temperature and hygrometry maintained at 20°C and 60-70%, respectively. After printing the slides were placed inside a slide storage box containing NaCl saturated solution, the box was sealed and incubation was performed at room temperature for a period of 36 hours (incubation can be performed for a minimum of 8 hours and a maximum of 80 hours). As a post spotting process the slides were washed twice for 5 minutes with ethanolamine to neutralize the epoxy groups that have not reacted (this step aims to avoid the binding of the excess of the probes or/and of the targets that could take part in the increase of the non specific background during the washing steps). The slides were then washed three times with deionised water to remove any remaining un-reacted ethanolamine. The slides were then immerse in passivation/washing solution (i.e. 1x SSC, 0.1% SDS (w/v) and 1% BSA (w/v) solution) using a slide rack and placed under gentle agitation at room temperature for 30 minutes. The slides were then rinsed with deionised water 3 x 5 minutes, and dried under a flow of nitrogen.

- **Assay description**

The library was always printed (by contact printing) in a 4 x4 format (16 spots per peptide) in different slides. The constructs were printed in PBS/DMSO (9:1 ratio). Subsequently the slides were placed in the Mic-Cell support chamber an into the LaVision scanner. Enzyme solutions with the number of units described in **table 5.6** were used in the amount of 0.5 mL in order to fill the chamber and completely cover the slide. The enzyme flow was constant at 1 μl per minute (some assays were carried out without any flow but with no success). All the described assays data are presented in annex and in all slides the pattern used for the printing had the position

pattern presented on **table 5.5** (A1 position being the one to the top right corner of the slide). Some tests were carried out using a smaller number of units per μl of enzyme but with no success. The assays for chymopapain and subtilisin were carried out at a controlled temperature of 25^{o}C while the thermolysin assays were carried out at 65^{o}C.

All the data treated in this assay was made using the software FIPS-2 from LaVision and is available in the support DVD with the data (the data for this section is available in DVD-2).

5.2.2.3. Microwells assays.

Microwell Fabrication[6]*:* Chitosan (Mw~300 kDa) and polystyrene (PS) (Mn~140,000 and Mw~230 kDa) were purchased from Aldrich. Prior to coating, the glass slides (obtained from Fisher Scientific, Silane-Prep^{TM}), were treated with oxygen plasma (Europlasma) for 20 minutes to clean the surface. These slides were then coated at 1500 rpm for 6s with 1% chitosan (in 2% aqueous acetic acid) using a spin coater (Model: P6708, Speedlines Technologies, IN, USA) and dried for sixteen hours under vacuum at 40°C. Subsequently the slides were coated with a solution of 10 % polystyrene in toluene (w/v). The coated slides were dried for 24 hours under vacuum at 40°C before contact printing. The microwell arrays were fabricated by contact printing (Genetix QArray mini; Hampshire, UK) with 16 aQu solid pins on the *dual-layer*-polymer coated slides at 15ºC. Multiple stampings of solvent (2:1 ratio of acetophenone/ ethyl acetate) was carried out with a delay of 200 ms between stampings. Solvents were then allowed to evaporate from the polymer surface for at least 48 hours at 40ºC before being used.

The library was printed (by contact printing, 10 times spot per well) in a 2 x 2 format (4 wells for each peptide; concentration of 25 μM). In the assays, the constructs were printed in glycerol/PBS/DMSO (4:4:1 respectively). The enzymes in known amounts were inkjet printed and the slides were scanned using a Nikon microscope equipped with the Pathfinder software (IMSTAR S.A., Paris, France). Enzyme solutions with the number of units described in **table 5.6** were inkjet printed in the amount of 20

drops per well. All the described assays data are presented in annex and in all slides the pattern used for the printing had the position pattern presented on **table 5.5** (A1 position being the one to the top right corner of the slide). To refer that the software to consult the data from IMSTAR is not freeware and therefore the data can only be viewed using a licensed Pathfinder version(IMSTAR S.A., Paris, France) (the data for this section is available in DVD-3).

5.2.2.4. Hydrophobic coated slide assays.

To SuperFrost glass slides (Menzel GmbH + Co KG, Saarbrückener Str. 248, D-38116 Braunschweig) 20 drops of 20% sucrose in water solution (w/v) were inkjet printed using a robotic printer Microdrop, and dried for 5 min at 80^{o}C (drops with d = 30 μm; 2 fields 16 x 16 with a distance of 0.9 mm; peptide concentration of 25 μM). The slides were then coated with (tridecafluoro-1, 1, 2, 2-tetrafluorooctyl) dimethylchlorosilane and washed with water (3 x) and acetone (3 x). Once dry, peptides in printing buffer and specific enzyme buffer (1:1) were printed (8 replicates; 20 drops d=30 μm) on the hydrophilic array previously generated, on both fields at 75 % humidity. Then, a 30% glycerol in water solution (w/w) was printed on top of the peptides to keep them in a semi-wet state. Assays were performed by printing the enzyme on the even positions (10 drops), while in the odd positions buffer was printed (10 drops) to provide an alternative control.

Arrays were then kept at 4^{o}C. Before enzymatic reactions were carried out, glycerol-based arrays were printed with enzyme buffer in order to swell spots. Then assays were done as described above.

The printing pattern used in the inkjet printing of the library differs from the ones used in the previous assays. Peptides were printed in the sequencial order described in **table 5.8**. These were printed using the printing pattern described in **figure 5.** in the amount of 8 sequential spots of the same substrate.

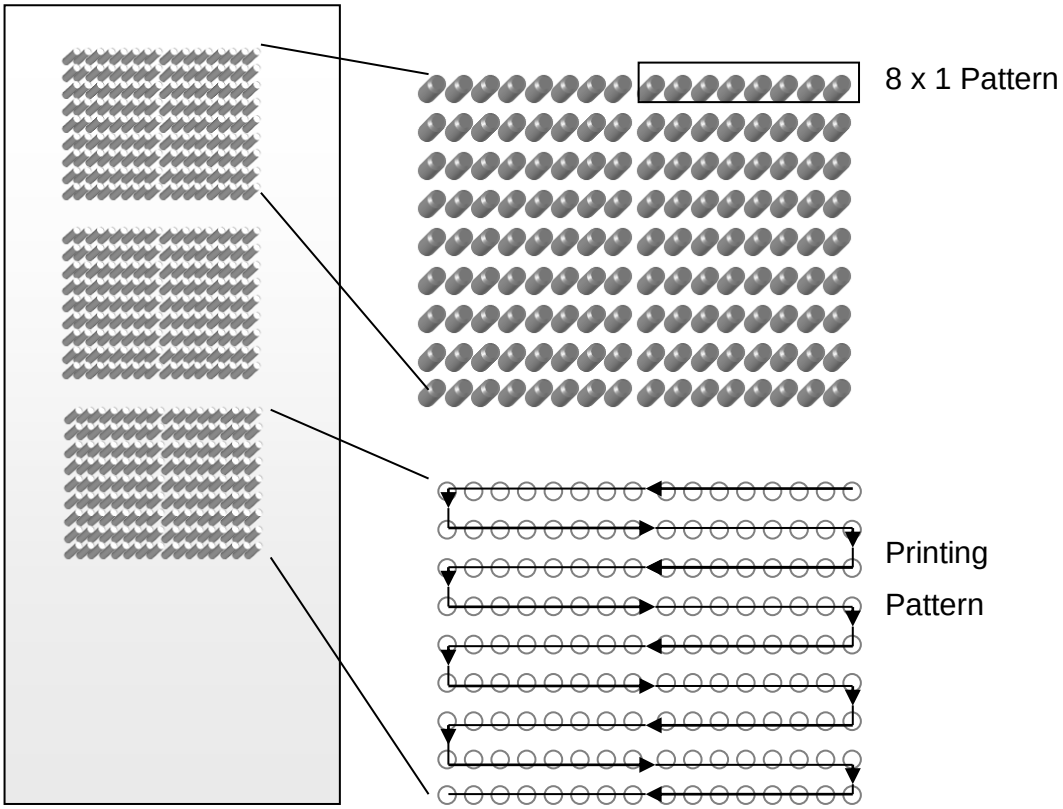


Figure 5.5 – printing pattern used in the hydrophobic coated slide.

Table 5.7 – Order of the peptides which were printed.

Sequences	Pos.	Sequences	Pos.	Sequences	Pos.	Sequences	Pos.
VNQHLK	1	YGGFLK	7	SALFIK	13	PIQQHK	19
EALYLK	2	GFFYYK	8	ALFIAK	14	HIPPYK	20
FFYTPK	3	EALYLK	9	SLFIAK	15	PIHYQK	21
ALYLVK	4	YQLENK	10	SAFIAK	16	HPIQQK	22
QHLCGK	5	YNQFFK	11	SALIAK	17	IHPQPK	23
LCGSHK	6	GSHLVK	12	SALFLK	18	PIPISK	24

Table 5.8 – Enzyme conditions used in the hydrophobic coated slide assays

Protease	Units per *µ*l	Buffer
Subtilisin	5×10^{-2}	0.1 M Tris pH 8.0; 5 mM $CaCl_2$
Chymopapain	1	0.2 M NaCl, 6. mM Mercaptoethanol; 5 mM EDTA; 50 mM *L*-Cysteine hydrochloride hydrate
Thermolysin	1×10^{-3}	0.1 M Tris pH 8.0; 10 mM $CaCl_2$

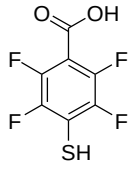
All the data treated in this assay was made using the software FIPS-2 from LaVision and is available in the support DVD with the data (the data for this section is available in DVD-4).

The treatment of the data was made using a dot diameter of 69 pixels with a tolerance of dot diameter of 1 pixel.

5.3. Experimental for Chapter Four.

5.3.1. Linker synthesis.

- **4-mercapto-2,3,5,6-tetrafluorobenzoic acid[7-9] (4.2).**



$C_7H_2F_4O_2S$

Mol. Wt.: 226.15

A solution of pentafluorobenzoic acid (8.5 g, 0.0403 mol) in aqueous sodium hydroxide (1.6g, 0.04mol in 22.5ml water) was added slowly to a stirred solution of sodium hydrosulphide (4g, 0.05mol) in aqueous sodium hydroxide (1.6 g, 0.04 mol in 7.5 mL water). The mixture was stirred at 70^{o}C for two hours, cooled and acidified (conc. HCl). The white precipitate was extracted into ether and the solvent was evaporated under reduced pressure, affording the crude product (8.1 g) which was crystallised from petroleum ether to give 4-mercapto-2,3,5,6-tetrafluorobenzoic acid (8 g; 0.0353 mol; 87% yield).

LRMS (ES^+) 226 (100%).

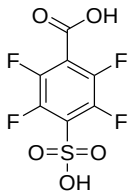
M.p. 151-152^{o}C.

FTIR (neat) υ_{max} = 3082, 1694, 1636, 1471 cm^{-1}.

^{13}C NMR (CD_3OD) δ 162.4, 146.1, 145.4, 118.0, 110.0.

^{19}F ($CDCl_3$) -138.0 (s, 2F), -140.1 (s, 2F).

- **4-sulpho-2,3,5,6-tetrafluorobenzoic acid[7-9] (4.3).**



$C_7H_2F_4O_5S$

Mol. Wt.: 274.15

Hydrogen peroxide (30%) (16 mL, 0.14 mol) was added to a stirred solution of 4-mercapto-2,3,5,6-tetrafluorobenzoic acid (8 g, 0.035 mol) in acetic acid (40 mL) at such a rate that the temperature was maintained between 65-70°C. When all the oxidant had been added, the mixture was heated for a further two hours at 70°C and then cooled to room temperature. Excess peroxide was destroyed by the addition of $K_2S_2O_5$. The water was removed under reduced pressure and the solid residue was dried to constant weight at 40°C/0.5 mbar pressure to give the 4-sulpho-2,3,5,6-tetrafluorobenzoic acid (7.7 g, 0.028 mol, 80% yield).

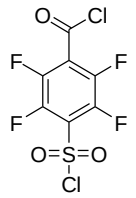
LRMS (ES^+) 274 (100%).

FTIR (neat) υ_{max} = 3500-2400, 3406, 3040, 1701, 1633, 1466 cm^{-1}.

^{13}C NMR (DMSO-d_6) δ 161.5, 145.2, 142.3, 129.0, 115.5.

^{19}F ($CDCl_3$) -140.6 (s, 2F), -138.3 (s, 2F)

- **4-chlorosulphonyl-2,3,5,6-tetrafuorobenzoyl chloride[7] (4.4).**



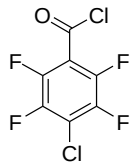
$C_7Cl_2F_4O_3S$

Mol. Wt.: 311.04

A stirred mixture of dry 4-sulpho-2,3,5,6-tetrafluorobenzoic acid (8.3 g, 0.0303 mol), phosporus pentachloride (18.8 g, 0.09 mol) and phosphorus trichloride (5 ml) was heated at 60^{o}C under nitrogen for two hours, when undissolved solid was removed by filtration. The filtrate was fractionally distilled (100°C, 0.2 mbar) to give the 4-chlorosulphonyl-2,3,5,6-tetrafluorobenzoyl chloride (4.9 g; 0.0159 mol; 64 % yield).

FTIR (neat) υ_{max} = 1794, 1765, 1485, 1401 cm^{-1}.
CIMS 275 (100%), 310 (11 %), 312 (9 %)
^{13}C NMR ($CDCl_3$) δ 157.6, 145.2, 141.7, 126.0, 122.5.
^{19}F ($CDCl_3$) -134.9 (s, 2F), -132.0 (s, 2F).

- **4-chloro-2,3,5,6-tetrafluorobenzoyl chloride[7].**



$C_7Cl_2F_4O$

Mol. Wt.: 246.97

4-sulpho-2,3,5,6-tetrafluorobenzoic acid sodium salt (2 g, 7.3 mmol) was covered with thionyl chloride (7.3 g, 62 mmol) and the suspension was stirred under $CaSO_4$ drying until gas evolution subsided. It was heated to reflux for 30 minutes, after which time DMF (0.15 g, 2 mmol) was added drop wise. The solution was heated in reflux for 2 hours. After the reaction was cooled, it was poured into ice. The resulting yellow solid was collected by filtration and dried in vacuum overnight. The crude material was characterized without further purification (1.6 g, 82% yield).

B.p. 101-102^{o}C/0.2 mbar.

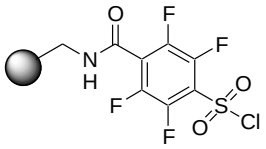
M.p. 58-60^{o}C

FTIR (neat) υ_{max} =3622, 3483, 1800, 1701, 1637, 1480, 1402, 1297, 1229, 1097, 973, 858 cm^{-1}

^{13}C NMR in DMSO δ_C = 159.7, 145.5, 142.2, 113.9, 112.6 ppm,

^{19}F NMR in DMSO δ_F = -140.0, -140.8 ppm.

- **2,3,5,6-tetrafluoro-4-methylcarbamoyl-benzenesulfonyl chloride polystyrene-supported[8,9] .**



To 1.0 g of aminomethyl polystyrene (1.10mmol g^{-1}) pre-swelled in anhydrous CH_2Cl_2 (20 mL), was added dropwise at 0ºC over 30 minutes with agitation a solution of 4-chlorosulfonyl-2,3,5,6-tetrafluorobenzoyl chloride in CH_2Cl_2 (0.34 g in 10 mL) and agitation was continued for one hour under nitrogen. *N*,*N*-diisopropylethylamine (1.15 mmol, 0.20 mL) was added as a dilute solution in CH_2Cl_2 (10% v/v, 1.8 mL) dropwise with agitation over 30 minutes. The agitation was then continued for 12 hours. The resin was filtered and washed with anhydrous CH_2Cl_2 (3×20 mL), and diethyl ether (3×20 mL) under nitrogen, then dried *in vacuo* at 40^{o}C over 12 hours to give the title compound.

FTIR (neat) υ_{max} = 1690, 989 cm^{-1}.

5.3.2. Qualitative test of the immobilized sulfonyl chloride.

Sulfonyl chloride resin was shaken with a solution of ethylenediamine in DMF (5% v/v) for 5 minutes at room temperature then washed thoroughly with DMF, THF, and Et_2O. The resin was then shaken with a solution of bromophenol blue in *N*,*N*-dimethylacetamide (1% w/v) at room temperature for 1 minute. The resin was washed thoroughly with DMF, DCM, and Et_2O. Persistent dark blue colouration of the resin indicated reactivity of the sulfonyl chloride resin with the diamine.

5.3.3. Stability studies of the tetrafluoroarylsulfonyl chloride linker.

Commercial suppliers of sulfonyl chloride substituted polystyrene recommend storage of the resin under dry and inert conditions, and preferably at low temperature over long time periods. In order to assess the storage properties and long-term stability of the tetrafluoroarylsulfonyl chloride substituted AMP resin **4.5**, a control test was made over a period of 3 months. A batch of freshly prepared immobilised tetrafluorosulfonyl chloride **4.5** was washed and dried thoroughly prior to division into four portions each treated as follows:

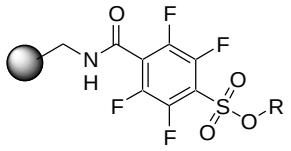
- Batch 1 was reacted with 7-hydroxy-4methylcoumarin under the normal conditions for sulfonate formation; the resulting resin was washed and dried thoroughly and stored at room temperature under air in a selected container for three months.
- Batch 2 of the sulfonyl chloride resin was stored at room temperature under air in a loosely sealed container for three months.
- Batch 3 of the sulfonyl chloride resin was stored at room temperature under argon in a sealed container for three months.
- Batch 4 of the sulfonyl chloride resin was stored at approximately -4 ^{o}C under argon in a sealed container for three months.

Following the three month storage period, samples of batches 2, 3 and 4 were reacted with 7-hydroxy-4methylcoumarin under the standard reaction conditions, washed and dried thoroughly.

All batches of the sulfonate derivatised resin were then subjected individually to the standard conditions developed for reductive cleavage, and a comparison of product yield was made to assess reactivity of the differently treated starting sulfonyl chloride resin batches. After isolation of the coumarin cleaveage product, it was shown that as expected, the tetrafluorosulfonyl chloride resin is susceptible to atmospheric decomposition, and batch 2 of the resin produced only traces of the desired product upon cleavage. Resin batches 1, 3 and 4 upon cleavage produced

essentially identical results, which surprisingly showed that as long as the supported sulfonyl chloride is stored under moisture free and inert conditions, its functionality is unaffected over a period of three months storage. This result also serves to prove the long-term stability of the derivatised sulfonate ester (batch 1 was stored for 3 months prior to cleavage), and pleasingly shows that no special storage conditions are required once starting sulfonyl chloride is derivatised.

5.3.4. Preparation of immobilised sulfonates.

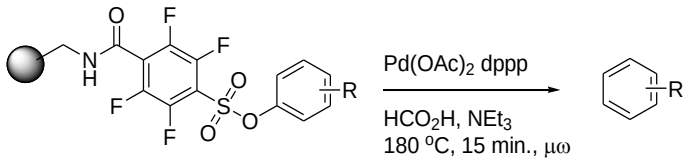


To slurry of freshly prepared polystyrene-supported sulfonyl chloride (1g, typical loading 0.75 mmol/g) pre-swollen in DCM (20 mL/g) under nitrogen, was added a mixture of the phenol or alcohol (3mol equiv., 2.25 mmol) and DIPEA (3 mol equiv., 2.25 mmol, 0.44 mL). The reaction mixture then shaken under nitrogen for 18 hours at room temperature, the resin was filtered and washed thoroughly with DMF, THF, DCM and Et_2O (3x 20 mL each) and dried *in vacuum* for 12 hours to give the resin bound sulfonate (loading results for the samples prepared were presented in **table 4.1**; section **4.2**).

Elemental analysis for compound **4.13.**

ELEMENT	**S**	**Br**
% Found 1	0.77	2.98
% Found 2	0.75	3.17

5.3.5. Representative procedure for reductive cleavage of the sulfonate linker under microwave irradiation.



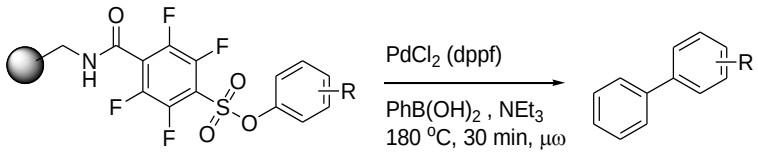
To aryl sulfonate resin (0.50g, 0.75 mmolg^{-1} typical sulfonate loading) swollen in DMF[a] (5.0 mL) was added $Pd(OAc)_2$ (10 mol %, 0.037 mmol, 8.5mg), 1,3-bis(diphenylphosphino)propane (15 mol %, 0.06 mmol, 22.5 mg) and a mixture of HCO_2H (9 mol equiv, 6.75 mmol, 128 μL) and NEt_3 (6.5 mol equiv, 4.875 mmol, 0.34 mL)[b]. The mixture was submitted to microwave irradiation for 15 minutes, at 180^{o}C, filtered, and the resin washed alternately with Et_2O and THF (5x 10 mL each washing cycle). Combined washings were gently concentrated *in vacuo* to remove Et_2O, THF and NEt_3. To the remaining DMF mixture was added an equal volume of water, and the mixture extracted with Et_2O (6x 10 mL). Combined ether extracts were concentrated *in vacuo* to a small volume (2-4 mL), and eluted through a short silica pad. Appropriate fractions were collected and concentrated *in vacuo*, and purified by preparative TLC (1 mm SiO_2), eluting with 5% v/v ethyl acetate/hexanes. Appropriate regions were removed, the silica extracted with ethyl acetate (6x 10 mL), and combined extracts were concentrated *in vacuo* to give the desired deoxygenated products.

All obtained products in Table 4.2 (section **4.4**) were characterized.

[a] Anhydrous *N,N*-dimethylformamide was degassed with nitrogen over 2 hours at room temperature immediately prior to its use.

[b] A mixture of triethylammonium formate in formic acid was prepared by dropwise addition with cooling, of 98 % formic acid to triethylamine.

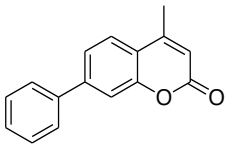
5.3.6. Representative procedure for the Suzuki cross-coupling of the sulfonate linker under microwave irradiation.



The arylsulfonate resin (0.50g, 0.75 mmolg^{-1} typical sulfonate loading) swollen in DMF[a] (5.00 mL) was added $PdCl_2$ (dppf) (10 mol %, 0.037 mmol, 32.5 mg), phenylboronic acid (3 mol equiv, 1.13 mmol, 137.0 mg) and NEt_3 (6 mol equiv, 2.3 mmol, 315 μL). The mixture was submitted to microwave irradiation for 30 minutes with agitation, at 180°C, filtered, and the resins washed alternately with Et_2O and DCM (5x 10 mL each washing cycle). Combined washings were gently concentrated *in vacuo* to remove Et_2O, DCM and NEt_3. To the remaining DMF mixture was added an equal volume of water, and the mixture extracted with Et_2O (6x 10 mL). Combined ether extracts were concentrated *in vacuo* to a small volume (1-2 mL), and eluted through a short silica pad. Appropriate fraction was collected and concentrated *in vacuo*, and purified by preparative TLC (1 mm SiO_2), eluting with 5% v/v ethyl acetate/hexanes. Appropriate regions were removed, the silica extracted with ethyl acetate (6x 10 mL), and combined extracts were concentrated *in vacuo* to give the desired product.

[a] Anhydrous *N,N*-dimethylformamide was degassed with nitrogen over 2 hours at room temperature immediately prior to its use.

- **4-methyl-7-phenyl-chromen-2-one.**



$C_{16}H_{12}O_2$

Mol. Wt.: 236.27

Obtained in 78% yield (1.0 mmol scale) as a colourless oil from the transfer deuteration of sulfonate immobilised 4-hydroxy-benzophenone.

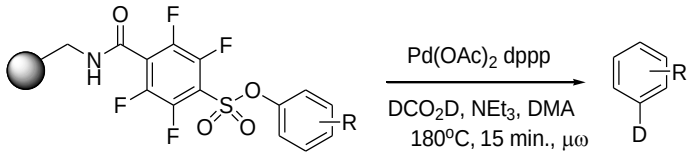
^{1}H NMR δ 7.80 (4H, d, *J* = 7.7 Hz), 7.58 (1H, dd, *J* = 7.35), 7.47 (4H, m).

^{13}C NMR δ 197.0, 137.8, 132.7, 130.3, 128.5, 128.4.

CIMS 183 (100%), 153 (34%), 105 (92%), 77 (82%), 51 (60%).

FTIR (neat) ν_{max} = 3059, 1657, 1275 cm^{-1}.

5.3.7. Representative procedure for deuterative cleavage of the sulfonate linker.



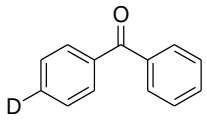
To arylsulfonate resin (0.5 g, 0.75 mmol/g^{-1} typical sulfonate loading) swollen in DMA[a] (4.1 mL) was added $Pd(OAc)_2$ (10 mol%, 0.075 mmol, 8.5 mg), 1,3-bis(diphenylphosphino)propane (15 mol %, 0.11 mmol, 22.5 mg) and a mixture of DCO_2D (18 mol equiv, 13.5 mmol, 205 μL) and NEt_3 (13 mol equiv, 9.75 mmol, 0.68 mL).[b] The mixture was submitted to microwave irradiation at 180^{o}C for 15 minutes, filtered, and the resin washed alternately with Et_2O and THF[c]. Combined washings were gently concentrated *in vacuo* to remove Et_2O, THF, and excess NEt_3. To the remaining DMA mixture was added an equal volume of water, and the mixture was extracted with Et_2O (6 x 10 mL). Combined ether extracts were concentrated *in vacuo* to a small volume (2-4 mL), and eluted (Et_2O) through a short silica pad[c]. Appropriate fractions were collected and concentrated *in vacuo*, and purified by preparative TLC (1 mm SiO_2), eluting with 5% v/v ethyl acetate/hexanes. Appropriate regions were removed, the silica extracted with ethyl acetate (6 x 10 mL), and combined extracts were concentrated *in vacuo* to give the deuterated products (72-84 % yield).

[a] Anhydrous *N*,*N*-dimethylacetamide was degassed by sparging with nitrogen over 2 hours at room temperature immediately prior to its use.

[b] A mixture of triethylammonium deuterate in deuterated formic acid was prepared by dropwise addition with cooling, of 99% deuterated formic acid to triethylamine (freshly distilled from CaH_2).

- [c] Careful washing of the reacted resin is required to achieve the tabulated yields of desired products; in all examples, 5 washing cycles each of sufficient volume to slurry the resin was sufficient.

- **(4-deuterio-phenyl)-phenyl-methanone (4.38).**



$C_{13}H_9DO$

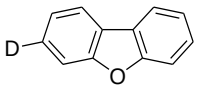
Mol. Wt.: 183.22

Obtained in 81% yield (1.0 mmol scale) as a colourless oil from the transfer deuteration of sulfonate immobilised 4-hydroxy-benzophenone.

^{1}H NMR δ 7.80 (4H, d, *J* = 7.7 Hz), 7.58 (1H, dd, *J* = 7.35), 7.47 (4H, m).
^{13}C NMR δ 197.0, 137.8, 132.7, 130.3, 128.5, 128.4.
CIMS 183 (100%), 153 (34%), 105 (92%), 77 (82%), 51 (60%).
FTIR (neat) ν_{max} = 3059, 1657, 1275 cm^{-1}.

These data were similar to those reported in the literature.[7]

- **3-deuterio-dibenzofuran (4.36).**



$C_{12}H_7DO$

Mol. Wt.: 169.20

Obtained in 74% yield (1.0 mmol scale) as a colourless crystalline solid from the transfer deuteration of sulfonate immobilised 2-hydroxy-dibenzofuran.

M.p 83-84^{o}C (Literature value 83-84^{o}C).

^{1}H NMR δ 7.97 (2H, m), 7.60 (2H, d, *J* = 8.3 Hz), 7.47 (2H, m), 7.36 (1H, dd, *J* = 7.4 Hz).

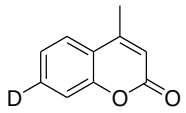
^{13}C NMR δ 156.5, 127.5, 127.3, 124.6, 123.0, 121.0, 120.9, 112.0.

CIMS 169 (100%), 140 (73%), 114 (36%).

FTIR (neat) ν_{max} = 3045, 1592, 1439, 1189 cm^{-1}.

These data were similar to those reported in the literature.[7]

- **7-deuterio-4-methyl-chromen-2-one (4.37).**



$C_{10}H_7DO_2$

Mol. Wt.: 161.18

Obtained in 80% yield (1.0 mmol scale) as a colourless crystalline solid from the transfer deuteration of sulfonate immobilised 7-hydroxy-4-methyl-coumarin.

M.p 91-93^{o}C (Literature value 92-94^{o}C).

^{1}H NMR δ 7.60 (2H, d, *J* = 8.09 Hz), 7.31(1H, s), 6.26 (1H, s), 2.42(3H, s).

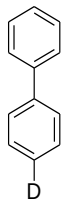
^{13}C NMR δ 160.9, 153.5, 152.6, 131.9, 131.6, 131.3, 124.7, 124.3, 120.0, 116.9, 115.1, 18.8, 18.5, 18.2.

CIMS 162 (96%), 134 (100%), 105 (28%).

FTIR (neat) ν_{max} = 3063, 1708, 1601, 1168 cm^{-1}.

These data were similar to those reported in the literature.[7]

- **4-deuterio-biphenyl (4.39).**



$C_{12}H_9D$

Mol. Wt.: 155.21

Obtained in 72% yield (1.0 mmol scale) as a colourless crystalline solid from the transfer deuteration of sulfonate immobilised 4-hydroxy-biphenyl.

M.p 69-71 ^{o}C.

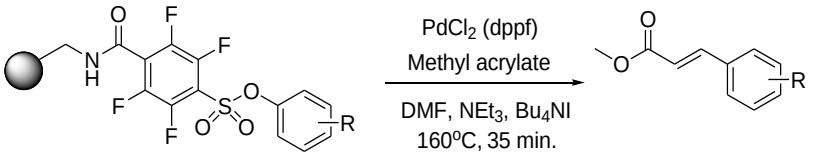
^{1}H NMR δ 7.65 (4H, d, *J* = 7.8 Hz), 7.49 (4H, m), 7.40 (1H, dd, J = 7.4 Hz).

^{13}C NMR δ 141.6, 129.1, 129.0, 127.6, 127.5.

CIMS 155 (100%), 127 (18%), 77 (23%).

FTIR (neat) ν_{max} = 3060, 3031, 1473, 1402 cm^{-1}.

5.3.8. Representative procedure for Heck cross-coupling of immobilised sulfonates with methyl acrylate.



To the arylsulfonate resin (0.5 g, 0.75 mmol/g^{-1} typical sulfonate loading) swollen in DMF[a] (4.35 mL) was added $PdCl_2$(dppf) (10 mol%, 0.075 mmol, 32.5 mg), methyl acrylate (10 mol equiv, 7.5 mmol, 0.34 mL), Bu_4NI (2 mol equiv, 1.5 mmol, 227.0 mg) and NEt_3 (6 mol equiv, 4.5 mmol, 0.315 mL). The mixture was shaken at 160°C for 35 minutes and the resin was filtered and washed alternately with Et_2O and CH_2Cl_2.[b] Combined washings were gently concentrated *in vacuo* to remove excess methyl acrylate and triethylamine. To the remaining DMF mixture was added an equal volume of water, and the mixture was extracted with Et_2O (6 x 10 mL). Combined ether extracts were concentrated *in vacuo* to a small volume (2-4 mL), and eluted (Et_2O) through a short silica pad. Appropriate fractions were collected and concentrated *in vacuo*, and purified by preparative TLC (1 mm SiO_2), eluting with 5% v/v ethyl acetate/hexanes. Appropriate regions were removed, the silica extracted with ethyl acetate (6 x 10 mL), and combined extracts were concentrated *in vacuo* to give the desired product (61-76% yield).

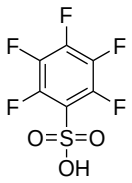
Spectroscopic data for products formed in these reactions were identical with those reported previously[8,9].

[a] Anhydrous *N*,*N*-dimethylformamide was degassed by sparging with nitrogen over 2 hours at room temperature immediately prior to its use.

[b] Careful washing of the reacted resin is required to achieve the tabulated yields of desired products; in all examples, 5 washing cycles each of sufficient volume to slurry the resin was sufficient.

5.3.9. Substrates for DoE screenings.

- **Pentafluorobenzenosulfonic acid (from pentafluorobenzene).**



$C_6HF_5O_3S$

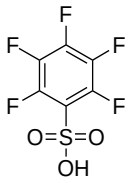
Mol. Wt.: 248.13

To a weighted quantity of pentafluorobenzene (2g) was dissolved in 1.14g of sulphur trioxide under reflux and dry conditions for 12 hours at 100^0C. The excess of sulphur trioxide was then neutralized by adding ice and barium carbonate. The formed barium sulphate was filtrated and the mixture was shaken with an ion exchange resin (AMBERLITE IR 120) for 3 hours. The solution is then filtered with activated charcoal and evaporated. The obtained product was then dried under reduced pressure at 100^0C for 8 hours to obtain the monohydrate. The monohydrate (1.5g) was then mixed with 3g of thionyl chloride for 8 hours at room temperature. The excess of thionyl chloride is then evaporated under reduced pressure at 60^0C (55% yield after purification).

CIMS (ES^-) 247.1 (100%),

^{19}F NMR in δ_F = -141,-152, -162 ppm.

- **Pentafluorobenzenosulfonic acid (from pentafluorobenzenethiol).**



$C_6HF_5O_3S$

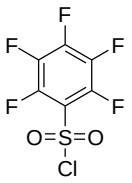
Mol. Wt.: 248.13

Hydrogen peroxide (30%) (16mL; 0.14mol) was added to a stirred solution of pentafluorobenzenethiol (5g; 0.0263 mol) in acetic acid (40 mL) at such a rate that the temperature was maintained between 65-70^{0}C. When all the oxidant had been added, the mixture was heated for a further two hours at 70^{0}C and then cooled to room temperature. The excess of peroxide was neutralized by addition of $K_2S_2O_5$ and most of the solvent was removed under reduced pressure. The resulting slurry was extracted with acetate from which a water-soluble fraction was then back extracted to water. The water was removed under reduced pressure and the solid residue was dried to constant weight at 40^{0}C/o.5 mbar to give pentafluorobenzenesulfonic acid (75% yield after purification).

CIMS (ES^-) 246.9 (100%).

^{19}F NMR in δ_F = -141,-152, -162 ppm.

- **Pentafluorobenzenosulfonyl chloride (from pentafluorobenzenesulfonic acid).**



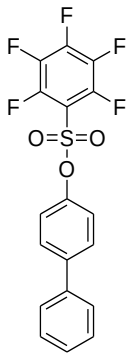
$C_6ClF_5O_2S$

Mol. Wt.: 266.57

A stirred mixture of dry pentafluorobenzenesulfonic acid (g, mmol), phosporus pentachloride (0.248g, 1mmol) and phosphorus trichloride (5 ml) was heated at 60^{o}C under nitrogen for two hours, then, the solid residue was removed by filtration. The filtrate was fractionally distilled under reduced pressure (100°C, 0.2 mbar) to give the pentafluorosulfonyl chloride. (0.170 g; 64 % yield after purification).

^{19}F ($CDCl_3$) = -139.17 (2F,m),-144.38 (2F,m), -162.06 (1F,m).

- **Synthesis of 4-(pentafluorobenzenesulfonato)biphenyl.**



$C_{18}H_9F_5O_3S$

Mol. Wt.: 400.32

2,3,4,5,6-Pentafluorobenzenesulfonyl chloride (0.56mL; 3.75 mmol) in dry dichloromethane (10 mL) was added drop wise to a solution of 4-phenylphenol (0.64g; 3.75 mmol) and triethylamine (1.05 mL; 9.40 mmol) in dry dichloromethane (30 mL). The reaction mixture was stirred at room temperature for 4 hours. The organic layer was then washed with 30 mL of water and the layers separated. The aqueous was extracted with dichloromethane (2 X 20 mL) and the combined organic extracts were dried with magnesium sulphate, filtered and concentrated under reduced pressure. The crude product was recrystallised from ethanol to afford the title compound (50 mg were isolated).

APCI (ES^-) 400 (M^+, 100%).

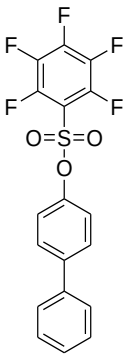
FTIR (neat) υ_{max} = 1651, 1382.

^{1}H NMR in $CDCl_3$= 7.61-7.38 (7H,m), 7.29-7.24 (2H,m).

^{13}C NMR in δ_C = 121.7, 127.1, 127.9, 128.8, 128.9, 139.4, 141.3, 143.2, 143.3, 147.7, 148.2 ppm.

^{19}F NMR in δ_F = -134, -143, -158 ppm.

- **4-(pentafluorobenzenesulfonato)biphenyl (5.1).**



$C_{18}H_9F_5O_3S$

Mol. Wt.: 400.32

A solution of the phenol (8.3mmol) in dry ether was added slowly to sodium hydride (0.4g; 10mmol as 60% dispersion in mineral oil, washed with ether) in dry ether (25 mL) at ice-bath temperature under stirring. Pentafluorobenzenesulfonyl chloride (2.22g; 1.357mL; 8.3mmol) was then added to the solution containing the phenoxide, cautiously to prevent the reaction from becoming exothermic. The reaction mixture was then refluxed for 3 hours, cooled, and an aqueous wash was performed. The organic layer was collected and the aqueous layer was washed once with ether. The combined organic layers were washed with a solution of sodium hydroxide (5% W/W), then dried with magnesium sulfate. The solvent was evaporated and the crude product characterised (75% yield after purification).

APCI (ES^-) 400 (M^+, 100%).

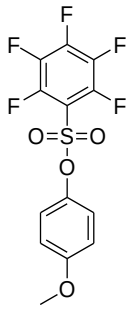
FTIR (neat) υ_{max} = 1651, 1382,

^{1}H NMR in $CDCl_3$= 7.61-7.38 (7H,m), 7.29-7.24 (2H,m);

^{13}C NMR in δ_C = 121.7, 127.1, 127.9, 128.8, 128.9, 139.4, 141.3, 143.2, 143.3, 147.7, 148.2 ppm,

^{19}F NMR in δ_F = -134, -143, -158 ppm.

- **1-methoxy-4-(pentafluorobenzenesulfonato)phenyl (5.2).**



$C_{13}H_7F_5O_4S$

Mol. Wt.: 354.25

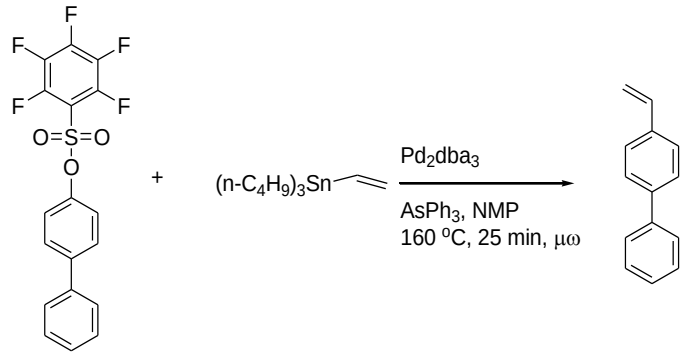
A solution of the phenoxide was prepared by adding (0.588 mL; 8.3mmol) of thalium ethoxide to 1.03g (8.3mmol) of 4-methoxyphenol dissolved in 25 mL of dry dichloromethane. This solution was then added drop wise to a solution of pentafluorobenzenesulfonyl chloride (1.357 mL; 8.3 mmol) in dry dichloromethane (25 mL). The addition was made keeping the solution at -20^0C for 1 hour. Once the addition was completed the solution was stirred for 1 hour at 0^0C and then stirred for a further 2 hours at room temperature. The solution was then washed with potassium carbonate solution (10% W/W; 2 X 50 mL), dried with magnesium sulphate and evaporated to obtain the desired product, which was then recrystallised from ethanol with a 31% yield after purification.

APCI (ES^-) 351.3 (100%).

FTIR (neat) υ_{max} =1655, 1384 cm^{-1}.

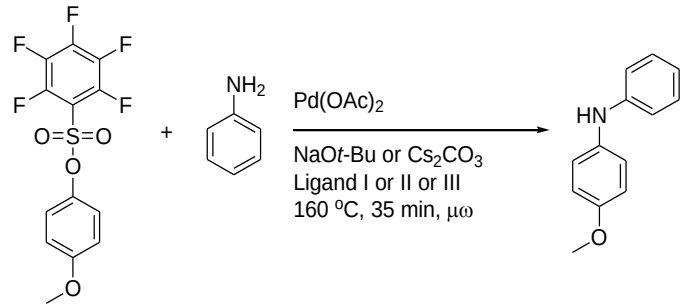
^{19}F NMR in δ_F = -133, -142, 158 ppm .

5.3.10. Representative procedures for the Stille cross-coupling of the sulfonate linker under microwave irradiation.



To **5.1** or **5.2** (1eq.) was added the catalyst (in 0.05 or 0.3 mol eq.), tributyl(vynil)tin (1.1 mol eq.) and ligand ($AsPPh_3$, PPh_3, or no ligand) (2 mol eq. to the amount of catalyst). The mixture was submitted to microwave irradiation for periods of time of 5, 10, 20 or 30 minutes with agitation, at temperatures ranging from 90 to 150^oC. A small sample was then characterized without workup, by HPLC using an ELSD and MS detector in order to determine the conversion of **5.1** or **5.2.** Solvents are presented in **table 5.12**, bases are presented in **table 5.10**, and catalysts in **table 5.9 (**tables presented at the end of this chapter).

5.3.11. Representative procedure for the Buchwald-Hartwig cross-coupling of the sulfonate linker under microwave irradiation.



To **5.1** or **5.2** (1eq.) was added the catalyst (in 0.05 or 0.3 mol eq.), aniline (1.1 mol eq.) and ligand (2 mol eq. to the amount of catalyst; see **table 5.11** where the used ligands are presented). The mixture was submitted to microwave irradiation for periods of time of 5, 10, 20 or 30 minutes with agitation, at temperatures ranging from 90 to 150^{o}C. A small sample was then characterized without workup, by HPLC using an ELSD and MS detector in order to determine the conversion of **5.1** or **5.2**. Solvents are presented in **table 5.12**, bases are presented in **table 5.10**, and catalysts in **table 5.9 (**tables presented at the end of this chapter).

Table 5.9 – Selected metals for the reactions under study.

	Quantity (g)	
Metal	**0.05 eq**	**0.3 eq**
M1 – Bis(triphenylphosphine) palladium (II) Chloride	0.00099	0.00594
M2 - Bis(triphenylphosphine) palladium (II) Acetate	0.00106	0.00634
M3 – Allyl Palladium Chloride dimmer	0.00052	0.00310
M5 – [1,1'-Bis(Diphenylphosphinoferrocene]dichloropalladium (II)	0.00103	0.00619
M13 – Bis(dibenzylideneacetone) Palladium(0)	0.00081	0.00486
M19 – Tris(dibenzylideneacetone) dipalladium(0)	0.00129	0.00775

Table 5.10 – Selected bases for the reactions under study.

	Quantity	
Bases	**2eq**	**6eq**
Sodium Hydroxide	0.002258 (g)	0.006774 (g)
Sodium acetate	0.00463 (g)	0.01389 (g)
Sodium carbonate	0.005982 (g)	0.017946 (g)
Cesium carbonate	0.0018389 (g)	0.055167 (g)
Triethylamine	0.00794 (mL)	0.02382(mL)

Table 5.11 – Selected ligands for the DoE of the Buchwald-Hartwig cross-coupling.

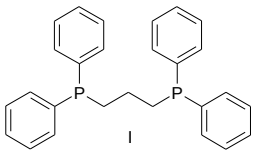
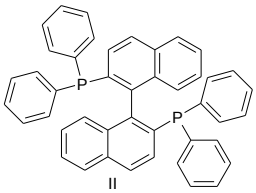
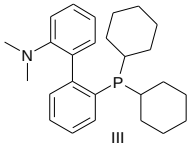
Ligands		
 I	 II	 III

Table 5.12 – Selected solvents for the DoE reactions under study.

Solvents	Description
NMP	*Nitrogen containing example*
DMF	*Nitrogen containing example*
MeCN	*Nitrogen containing example*
Dichloromethane	*Chlorinated example*
Methanol	*Alcohol example*
1,4-Dioxane	*Cyclic ether example*
Dibutylether	*Ether example*
THF	*Cyclic ether example*
Ethyl acetate	*Carbonyl example*
Heptane	*Hydrocarbon example*
Toluene	*Hydrocarbon example*
Xilenes	*Hydrocarbon example*
Water	*Protonic example*

5.4. References

(1) Sarin, V. K.; Kent, S. B. H.; Tam, J. P.; Merrifield, R. B. *Analytical Biochemistry* 1981, *117*, 147-157.

(2) King, D. S.; Fields, C. G.; Fields, G. B. *International Journal of Peptide and Protein Research* 1990, *36*, 255-266.

(3) Diaz-Mochon, J.; Bialy, L.; Bradley, M. *ORGANIC LETTERS* 2004, *6*, 1127-1129.

(4) Chhabra, S. R.; *et al.*, *Tetrahedron Letters* 1998, *39*, 1603 - 1606.

(5) Nguyen, T.; Francis, M. B. *Organic Letters* 2003, *5*, 3245 - 3248.

(6) Khan, F.; Zhang, R.; Unciti-Broceta, A.; Diaz-Mochon, J. J.; Bradley, M. *Advanced Materials* 2007, *19*, 3524-3529.

(7) Fielding, H. C.; Shirley, I. M. *Journal of Fluorine Chemistry* 1992, *59*, 15-31.

(8) Revell, J. D. PhD, Southampton, 2003.

(9) Revell, J. D.; Ganesan, A. *Chemical Communications* 2004, 1916-1917.